AQ

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	1997	Final Report

4. TITLE AND SUBTITLE

Study Into The Design Of The Semiconductor Power Pulser With Less Than 1 Nanosecond Front And Up To 100 KV Output

5. FUNDING NUMBERS

F6170896W0280

6. AUTHOR(S)

Prof. Alexei Kardo-Sysoev

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Ioffe Physico-Technical Institute
Ioffe Physico-Technical Institute
Academy of Science of Russia13Politechnicheskaya 26
St. Petersburg 194021
Russia

8. PERFORMING ORGANIZATION REPORT NUMBER

N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

EOARD
PSC 802 BOX 14
FPO 09499-0200

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

SPC 96-4074

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

A

13. ABSTRACT (Maximum 200 words)

This report results from a contract tasking Ioffe Physico-Technical Institute as follows: Study the design of semiconductor power pulsers with less than 1 nanosecond front and up to 100 kV output.

14. SUBJECT TERMS

Physics

15. NUMBER OF PAGES

120

16. PRICE CODE

N/A

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

SPC-96-4074

Study into the Design of the Semiconductor Power Pulsers with less than 1 nanosecond front and up to 100 kV output

Report

Part 1

Megapulse Ltd.

1997

1 A brief overview of the physical basis of the new solid state pulser design

The technology of short (less than 10^{-8} s) pulse generation is limited, as a rule, by availability of switching devices (closing or opening switches). The most promising devices are semiconductor devices, but until last years the power of semiconductor devices have been far lower, than the power of gas-discharge gaps or magnetic compressing cells.

In the last decade, a promising trend connected with the appearance of high-power, super fast semiconductor devices has arisen. The devices operation is based on recently discovered physical effects: superfast recovery of high voltage diodes, and delayed "overvoltaged" breakdown [1], [2].

1.1 New switching devices

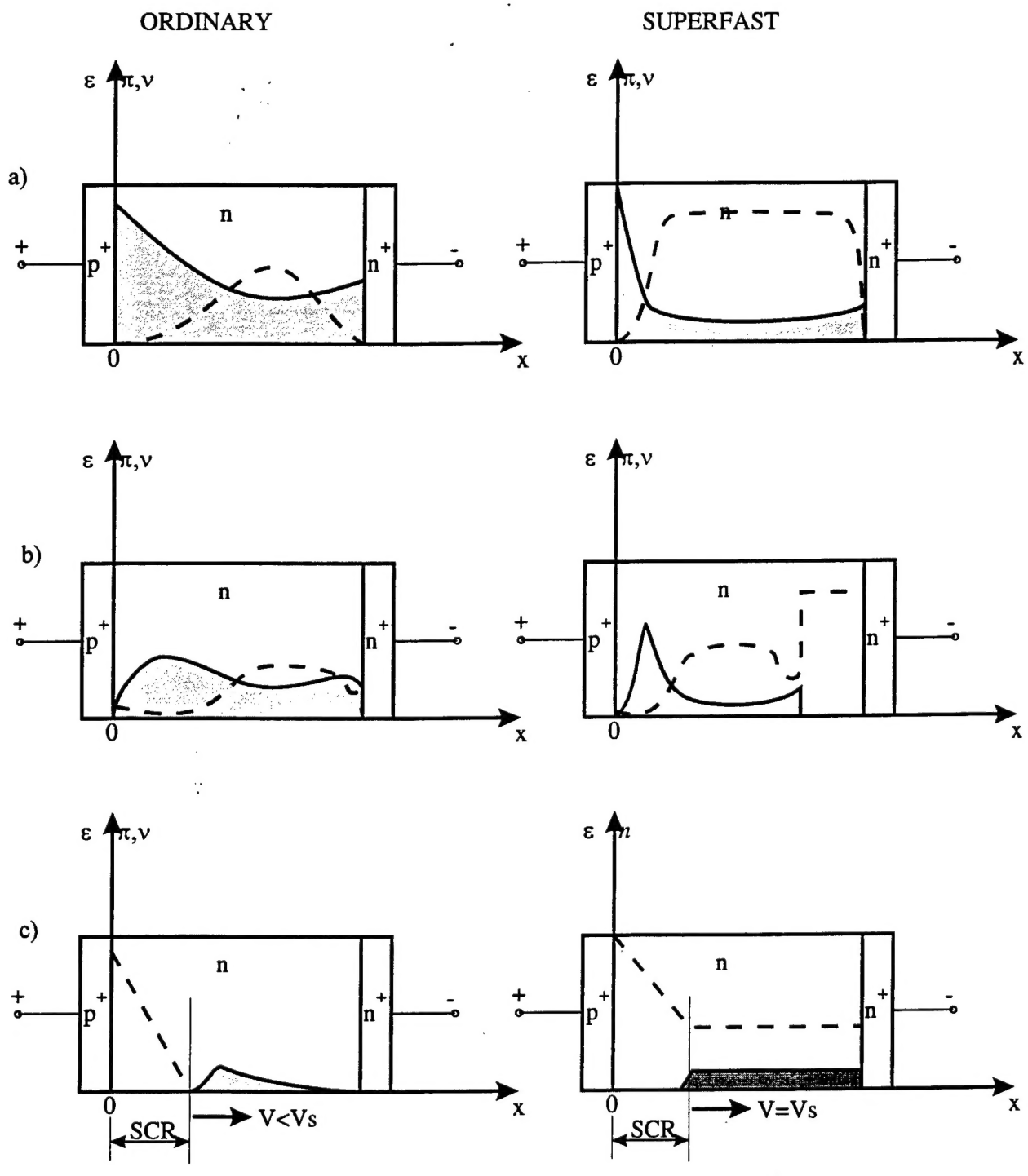
1.1.1 Nanosecond range

The problem of high power switching in nanosecond range can be solved by using the effect of superfast recovery of high voltage power diode when it is switched from forward to reverse bias.

The well known ordinary recovery process looks like following (Fig.1.1). The long forward current pulse produces relatively uniform high density electron-hole plasma distribution in the diode base layer (Fig. 1.1a). During the recovery process reverse current pulls out electron-hole plasma from the base. The plasma concentration at the p^+n-junction decreases to zero. Space Charge Region (SCR) occurs near the p^+n-junction and begins to move right (Fig. 1.1b). At this moment the diode resistivity begins to increase dramatically and the current decreases. This process is not fast because the plasma storage near the SCR boundary prevents the fast motion of this boundary (Fig.1.1c); this is a typical submicrosecond process.

But we have found the special conditions for two orders faster recovery (Fig. 1.1'). Very short (hundreds ns) forward current pulse forms steep not uniform plasma distribution in the diode base layer (Fig. 1.1'a). A thin high density ($\sim 10^{17} cm^{-3}$) plasma layer is formed just near the p^+n-junction by the diffusion process, the other part of distribution is formed by a fast bipolar drift wave and has two orders lower plasma concentration. Then the external voltage polarity is changed (Fig.1.1'b), SCR forms near p^+n-junction and it's boundary begins to move right slowly ($V<V_S$) as in an ordinary process. Simultaneously at the n+n- boundary the bipolar drift wave is formed and begins to move left; the front of this wave is very sharp

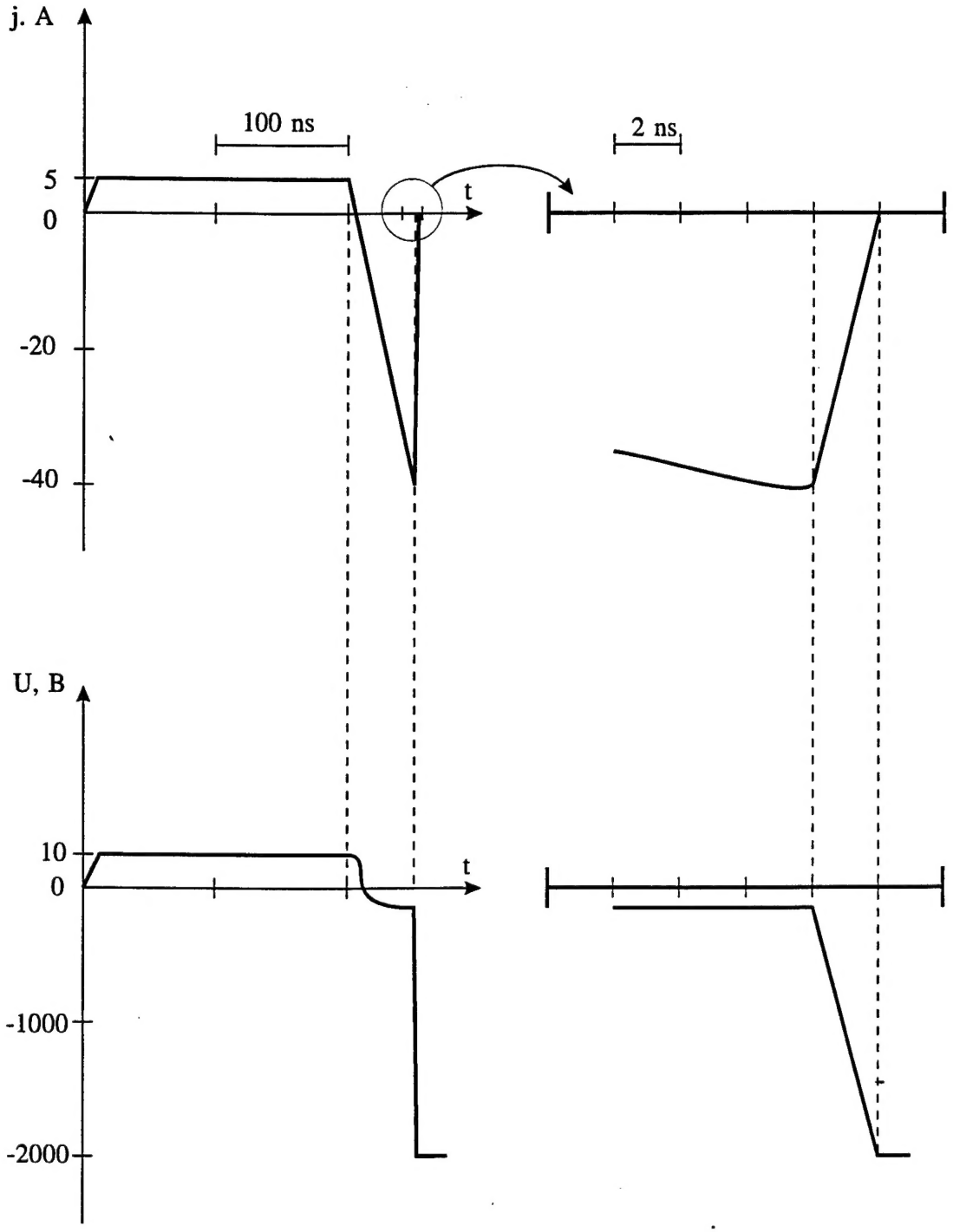
VOLTAGE RERCOVERY IN DIODES



-electric field distribution
-electron-hole plasma distribution
-majority elecron distribution
SCR - space charge region
V - velocity of SCR boundary
Vs - saturated velocity

Fig.1.1

Fig.1.1'



Transient process of superfast recovery in hihg voltage diodes

Fig.1.2

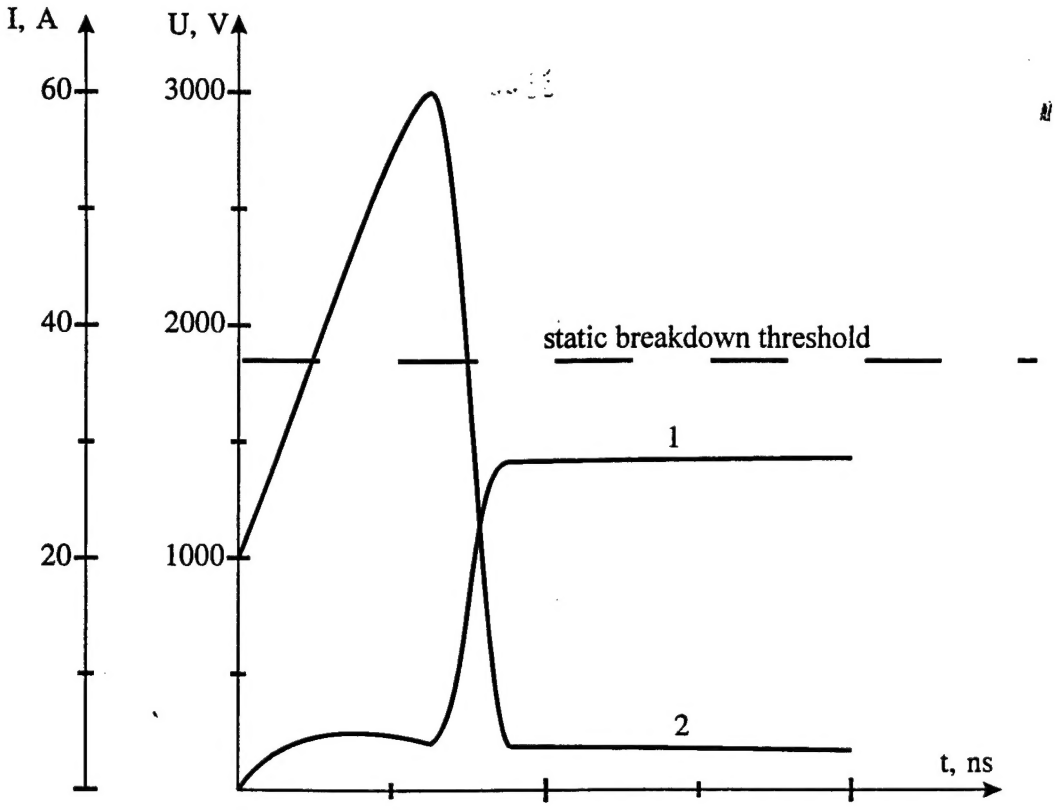
because of the front motion velocity is inversely proportional to the plasma concentration. Calculations show that this front arrives at the boundary of plasma layer exactly simultaneously with the complete depletion of this layer. Both of these processes are very fast: the first due to small dimension of plasma layer, the second - due to the high velocity of wave front motion. After this the SCR boundary begins to move right very fast because there is no plasma near the boundary and it's velocity depends on majority carriers motion only (Fig. 1.1'c). Correlation between reverse current density and base material resistivity should be chosen so that the electric field in the base layer should be high enough for carriers velocity saturation ($V=V_S$). If this is the case, the SCR boundary moves right under the saturated velocity Vs, the reverse voltage increases and reverse current decreases very fast. Typical value of voltage rise rate is 10^{12}V/S (1 kV per 1 ns) in the diode with 2 kV breakdown voltage. Fig.1.2 shows typical oscillograms of this process.

The diode for superfast switching is called a drift step-recovery diode (DSRD); it is an opening switch and can form the cut-off current front from 0.5 to 3.0 ns depending on the diode operating voltage (0.7 - 3 kV, correspondingly). The operating reverse current density depends on the operating voltage and for 2 kV it is equal to ~ 10^2A/cm^2. The operating area of such a device can be as high as 30-40 cm^2, so the current pulse may be as high as 3-4 kA.

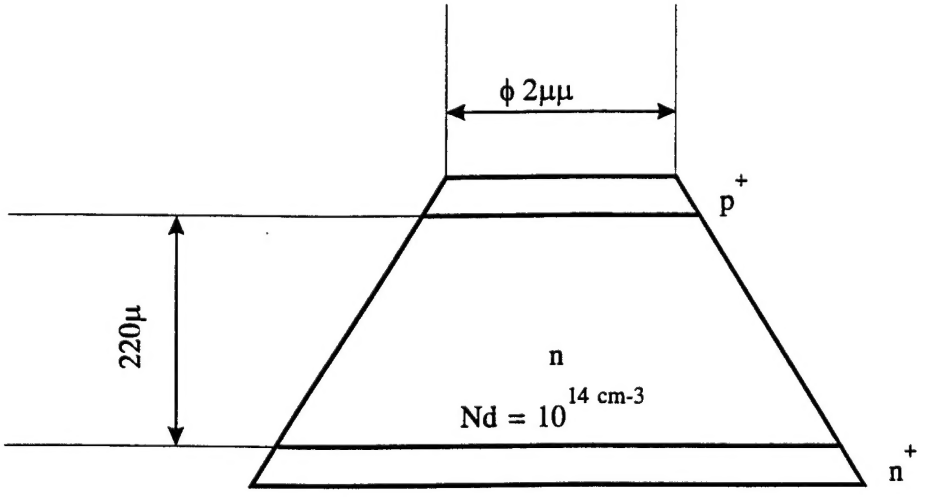
1.1.2 Subnanosecond and picosecond range

High power switching in the picosecond range cannot be based on the principles associated with the motion of carriers in semiconductors just as in micro and nanosecond ranges. Even when moving at saturated velocity the carriers travel in 100 ps only 10μ, which is about an order less than SCR width in power diode blocking, for instance, 2 kV. So the new physical phenomena are necessary for the picosecond switching. We have found one of such phenomena during the investigations of the silicon diode breakdown at superfast overvoltage.

Fig. 1.3 shows the results of these experiments. We applied to a reversly biased diode (the diode structure is shown on Fig. 1.3b) overvoltage pulse with rise rate more than 10^{12} V/sec, formed by DSRD-generator. One can see that during several nanoseconds the avalanche breakdown does not occur despite the applied voltage is two times higher than the breakdown voltage in static conditions. Then the voltage applied to the diode dramatically drops during tens of picoseconds and the current increases. The physical nature of this phenomenon was cleared up by detailed experiments . The breakdown delay comes from the fact (Fig. 1.4) that thermal generation cannot produce in a few ns a sufficient number of

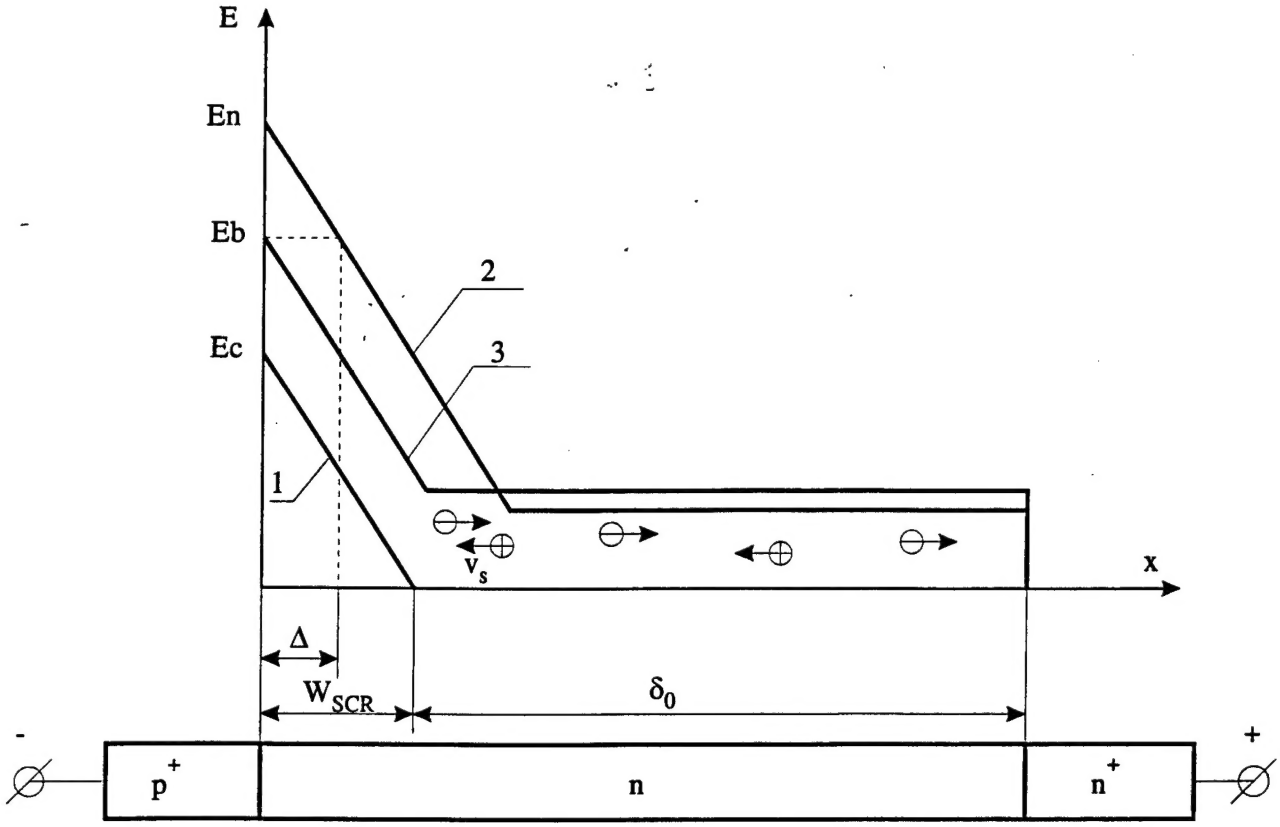


a) current (1) and voltage (2) curves during switching

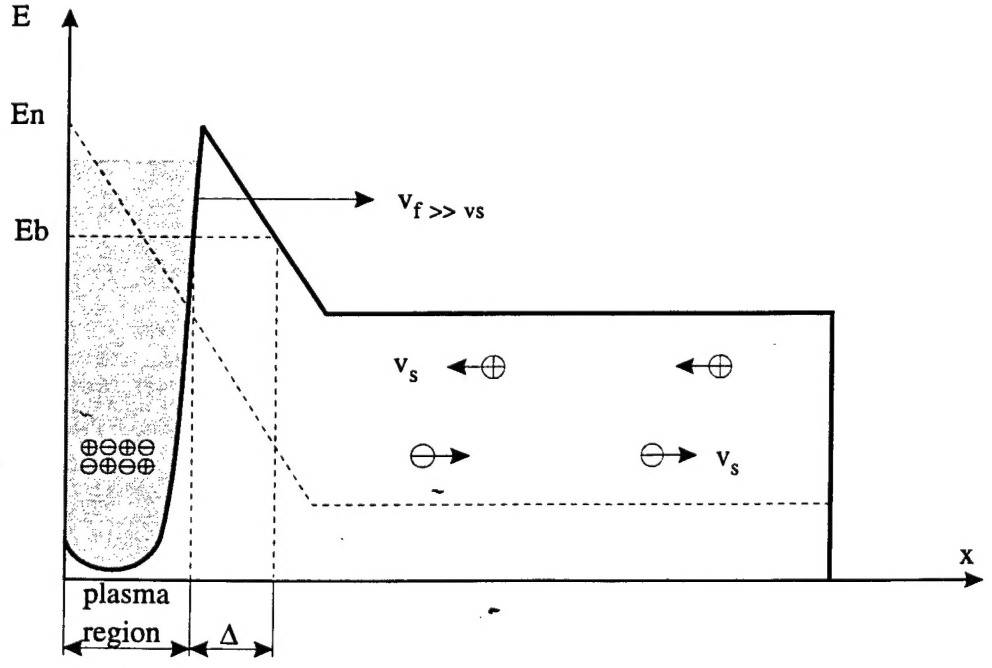


b) silicon diode structure under investigation

superfast reversible brakdown in high voltage diodes (delayed ionisatio)



a) Electric field distribution:
1- under static bias
2,3- under overvoltage
Ec - electric field maximum under static bias
Eb - breakdown electric field under static conditions
W_{SCR} - width of space charge region under static bias
Δ – ωιδτη οφ συπερηιγη φιελδ ρεγιον
δ_0 - width of diode base layer neutral region
vs - carriers saturated velocity



b) Impact ionisation wave propogation

carriers for lattice ionisation in superhigh field region. At the same time a high displacement current passing through the device produces in the neutral part of the base a field high enough for the lattice impact ionisation by majority carriers. The holes, produced here, drift left and for a few ns reach the superhigh field region near p^+n-junction. These holes initiate an extremely intensive breakdown in this region; the ionisation time here is less than 10^{-11} sec and at about this time the superhigh field region is filled up by electron-hole plasma and the field drops. In the nearby region the field increases and the breakdown begins, etc. The front of the so formed impact ionisation wave moves right toward the flux of the initiating breakdown holes. Current thorough the diode increases and after the wave front reaches n^+-region the base is filled by dense electron-hole plasma. The rate of this process is determined by the ionisation rate in the superhigh field region and can be of two orders higher than the saturated velocity of carriers. So the main point of this phenomenon is the space separation for a short time the superhigh field region and the holes generation region. Diodes based on this phenomenon are called silicon avalanche shapers, SAS. It is of interest to note that the same phenomenon occurs also in GaAs diodes and leads to the high power stimulated emission at the certain conditions. The shortest switching time obtained up to now in silicon diodes is 50 ps for 0.1 MW of switched power and jitter less than 30 ps.

1.2 Circuit engineering

1.2.1 Step recovery devices

Step recovery devices are used as fast opening switches in power nanosecond pulsers. However, as distinguished from traditional devices, for example transistors, step recovery devices have a time of conducting state less than a microsecond, usually near hundred nanoseconds. Therefore in real systems they are always used jointly with traditional switches: transistors or thyristors. Traditional switches generate initial relatively long pulses (hundreds nanoseconds), while step recovery devices compress them; they shorten fronts or/and decrease the lengths.

In order to force a drift step recovery device into the conducting state, a forward current, or "pumping" current is passed through it first. This current creates the necessary charge of excess (nonequlibrium) carriers. In diode the pumping current passes through the very same pair of electrodes as the current of the main (power) circuit does. Thus in systems using a two-electrode switches, the pumping circuit and the power circuit turn out to be connected. In some cases this connection may be used for transfer of energy out of the pumping circuit into the load, improving efficiency of pulse generation.

In other cases when it is desirable to weaken the connection, it is possible to separate pumping circuit from the load by use of highpass filters as a choke. Both of this possibilities have their advantages and disadvantages and will be examined later.

As was already noted, step recovery devices are current breakers. There are the following main variants of the pulse shaping systems:

--circuits with intermediate accumulation of energy, in particular in inductance (which may be made from inductive coil or a piece of transmission line). The output pulse is shaped during sharp break off of the current flowing through the inductance by step recovery devices (an example is shown at Fig. 1.5);

--circuits in which the device shorts the transmission line from the generator to the load for the time needed for establishing a wave with the required amplitude in the line. Then the device sharply "opens" the line and shapes pulse front (the case is shown at Fig. 1.6); the simplest variant is one when DSRD shunts the load directly.

A significant feature of all shaping circuits consists of the fact that during the time when the step recovery device is in the conducting state (τ_-) it is necessary to increase the current in energy storage inductor to the required level I_m, after which this current should be quickly cut off. The current increase is done by the primary switch. Powerful transistors (bipolar or field effect) or thyristors may be used as primary switches. In powerful pulse forming circuits with a large peak currents I_m, the primary switch should guarantee a relatively large current rise rate $I' = \frac{dI}{dt} \approx I_m/\tau_-$. It should be noted that the less is τ_-, the better switching properties of the step recovery devices are. Usually τ_- is 50-200 ns, that yields $I' \geq 1$ A/ns for $I_m > 100$ A. The problem of the high dI/dt capability of power closing switches will be examined later in detailed way.

Although schematic of step recovery devices (fast opening switches) strongly differs from those for widely used schematic of closing switches, it is possible to design power switch combined from slow closing thyristor and fast opening switches. Such thyristor-diode closing switch (TDCS), shown in Fig. 1.7, may be considered as fast closing two terminal switch (1 and 2 at Fig. 1.7). In this case, it is possible to use the great experience in the development of powerful thyristor-based pulsers, after substitution of thyristor by TDCS with dozens-fold improvement in turn-on rise time.

In the next chapters the most efficient DSRD-based circuits will be considered.

a. Quasi -symmetrical parallel LC circuit [3]

Let us consider two similar LC- circuits with DSRD inserted in common branch of the circuits (Fig. 1.5). Initially energy storage capacitors C1, C2 are charged, (the polarity is shown on Fig. 1.5) and switches S1, S2 are opened. When the switch S1 is closed, capacitor C1 discharges via inductor L1 and the diode (DSRD). The discharge current is forward current for DSRD, the resistance of DSRD is low and the current (I1) in C1, S1, L1, DSRD circuit oscillates. Half-cycle period of LC circuit must not be more than several hundreds nanoseconds. Minority carriers lifetime in DSRD is large - ten's and more microseconds. The total amount of electron-hole pairs, injected ("pumped") in DSRD during the first forward halfperiod of current oscillation is equal to the charge passed through the diode.

When the current changes it's direction (the second halfperiod) the diode remains in high conducting state due to stored electron-hole pairs. If at the moment, when current I1 crosses the zero -level ($t = \tau_-$), the second switch S2 closes, the C2 discharge current (I2) is added to the current of L1, C1 circuit. The total DSRD's current doubles. At the moment of the current maximum extracted for τ_- period of time from DSRD charge is equal to injected during τ_- period and DSRD total current sharply (for 1-2 nanoseconds) is cut. At the moment all energy, initially stored in C1, C2 capacitors, are accumulated in L1, L2 inductors, and theirs currents reach maximum values. During the time of DSRD switch off processes its current is switched into the load resistor Rl. The front of the load current pulse is determined by turn off time of the DSRD, and the decay is ~ L/Rl (where L is the total inductance of L1 and L2 connected in parallel). The peak load voltage Ulm ~ Rl.(I1+I2) may be many times higher (> 10 times) than the initial capacitor voltage Vo. Such voltage multiplication is one of the important advantages of circuits based on DSRD. High voltage exists in circuit only very short time - nanoseconds, therefore high voltage corona and arc discharges troubles are not so severe as in case of DC.

The scheme of case of the ideal primary switches S1, S2 shown above is very simple and very effective.

It should be mentioned, that total delay of shaped pulse is equal to 3/4 of LC circuit period and may reach hundreds nanosecond. The stability of the delay is determined by stability of LC circuit, which, as well known, may be better than 10^{-4} , so jitter of the delay may be less than 100 ps. When C1 and C2 are charged from one power supply, unstability of the supply does not influence the moment of DSRD current brake (the ratio of injecting current to extracting current does not depend on power supply voltage).

Unideality of parts determined by loss of energy in L, C, the primary switches S1 and S2 may strongly influence on the efficiency of pulse generation.

Influence of losses may be defined by Q-factor of C,S,L, DSRD circuit or by the ratio of first current maximum at the first half-cycle to the second, when only one switch S1 is turned on. It turned out that the task of getting good Q-factor (> 3) for high currents > 100 A is very hard.

It is very easy to get large Q-factor (> 100) for small (< 1 microhenry) ironless linear inductors even at high (> 1000A) currents. Q-factor of low inductive capacitors with stored energy D> 10^{-3}J is not so large as for inductors and for ceramic best capacitors is ~ 10-30.

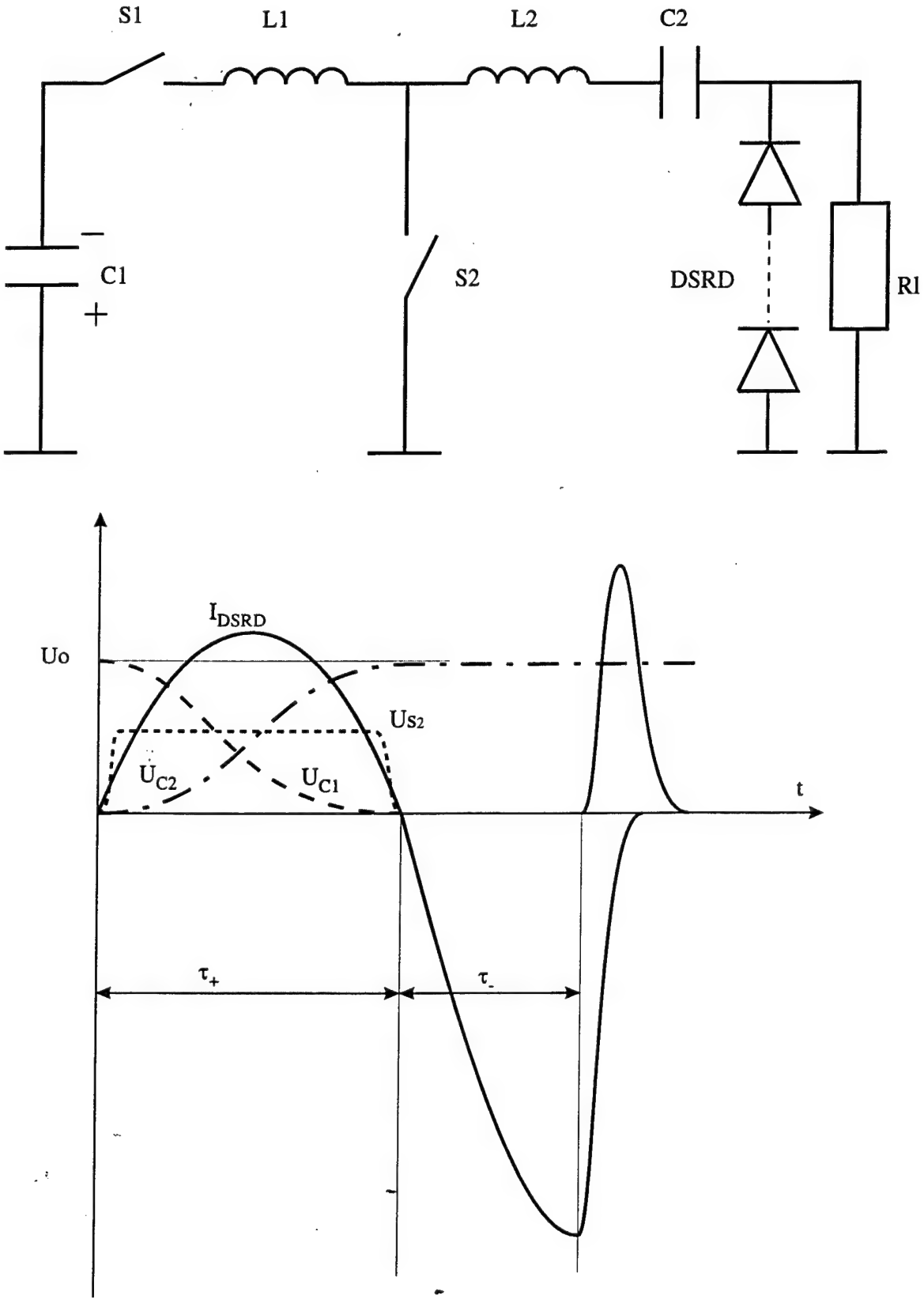
The losses in switches play major role on Q-factor. There are two types of losses: transient losses - during turn-on processes and losses due to resistance in turn-on state. Actually they decreases Q factor of LC circuits down to the factor of two-five.

Due to low Q-factor symmetry of LC circuit (Fig. 1.5) is broken. With equal L and C in both circuits, the second halfperiod L1, C1 current I1 is not equal to the current I2 during the first halfperiod L2, C2 . When the DSRD breaks current, part of L2 current (I2) goes to the first inductor L1 additionally increasing total losses. Adjustment of parameters of the first LC circuit (increase of L1 and decrease of C1) can decrease the current interchange between L1 and L2, but, for example, if Q-factor is less than 3, overall efficiency of pulse forming circuit still is worse than 70-80%.

Every kind of known closing switch may be used as a primary switch S1 and S2. In this work, later we will consider properties of only semiconductor devices for use as a primary switches: transistors (fields effect and bipolar), thyristors and dynistors. In this circuit current in both arms are virtually independent.

b. Series LCs circuit

The other example of circuits is shown in Fig.1.8, where interchange of inductors currents plays major role in series LC's circuit. Initially switches S1, S2 are opened, the storage capacitor C1 is charged up to initial voltage Uo, the second capacitor C2 is discharged. After closing of the first switch S1, the first capacitor C1 discharges, the second C2 charges. If capacitances of C1 and C2 are equal after halfperiod time of LC circuit cycle, the second capacitor C2 will be charged up to the initial voltage Uo. The discharge current flows through DSRD in forward direction and "pumps" it as in the previous case. At the moment of maximum voltage at C2 the second switch S2 closes and C2 discharges via S2, L2 and DSRD (current flows in blocking direction for DSRD, but as in previous case, DSRD is



series LC circuit for DSRD

Fig.1.8

in conducting state). If L1 = L2, C1 = C2, the impedance of L2, C2 circuit is two times less than the impedance of C1, L1, L2, C2 circuit at the "pumping" stage and the peak current is two times more than pumping current. Therefore at the moment of current maximum the extracted charge is equal to pumped charge. The DSRD breaks current and current flows into the load Rl.

If Q-factor is small, but voltage drop on S2 in open state is low, L1 and L2 are separated still and the decay of load pulse is determined by L2 only.

The new features of the series LCs circuit are:

a) the first switch has to pass all energy;

b) the second switch has to hold total load current;

c) after closing of the first switch, the sharp voltage pulse appears at the second switch S2 (Fig. 1.8). The last feature can lead to the stray dU/dt turning on of the S2, if S2 is a thyristor, but in case of magnetic switch S2 the voltage may be used to turn on magnetic switch [4]. The first and the second features (a, b) are disadvantages in the case of a high current (high energy pulse generation) to compare with the parallel LCs circuit. A piece of transmission line (coaxial cable) may be connected instead of inductor L2. In this case a rectangular pulse will be generated and pulse length equals doubled wave propagation time along the line.

The same substitution of lines instead of inductors in parallel circuit (Fig. 1.5) permits to shape rectangular pulses, but in this case current exchange between two parts of circuit could distort the ideal shape.

c. Thyristor-diode closing switch (TDCS)

The circuit diagram for switching on the DSRD- thyristor pair given in Fig. 1.7 makes it possible to use this pair, together with storage capacitor C, in the traditional circuit technology as a combined closing switch with a switching time in the ones of nanoseconds. In this case, the thyristor Q is connected in the usual manner to the linear modulator system with the forming line (C1-C3, L1-L2). An additional circuit consisting of the DSRD, capacitor Cp and the pumping source is connected in parallel to thyristor Q. Initially the DSRD is in conducting state due to previous pumping. When the thyristor is switched on , the current through it increases. Nevertheless, the initial voltage Vo, which is equal to the voltage of the charge of the forming line, is retained on it due to capacitor Cp. In the process of increasing of the current, carriers are accumulated in the thyristor, and its conductivity is modulated. The thyristor can be represented in the form of resistance $R(t) = V_0/I(t)$ which changes in time.

Since the voltage on the thyristor does not change, current does not arise in the modulator load (the voltage on the forming line is precisely compensated for by voltage V_0).

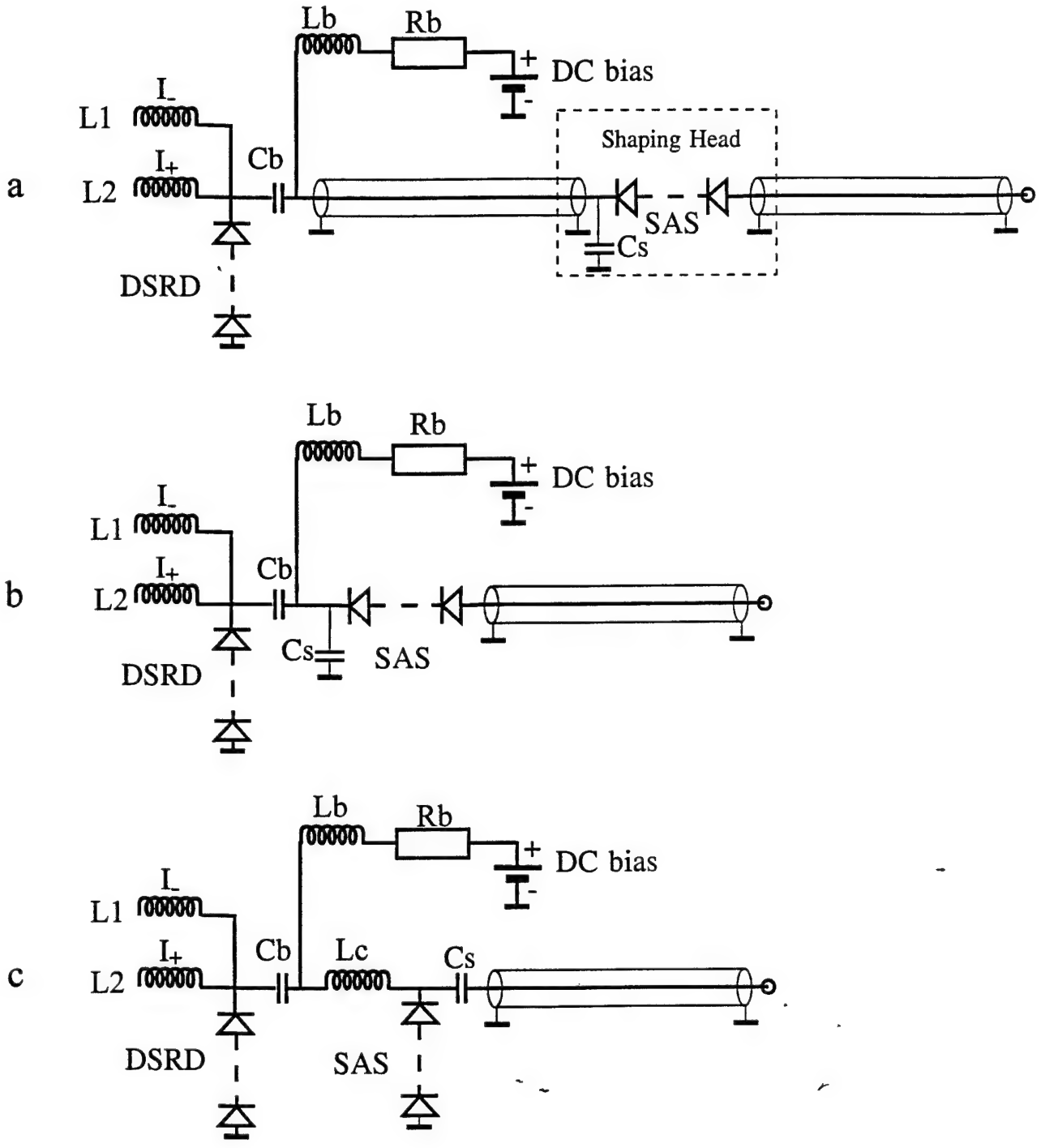
At moment $t = \tau_-$ when charge stored in the DSRD (Q_s,) is pulled out and the thyristor current reaches the value $I_m \approx Q_s/\tau_-$ the current through the DSRD is broken (the duration of the break is τ_f), and the capacitor Cp is disconnected from the thyristor. At this moment the thyristor has dynamic resistance $R_- = V_0/I_m$. The voltage of the forming line V_0 is applied to the series connected thyristor and load resistance Rl, in which there arises a current step with amplitude $dV_0/(2R_H + R_-)$ and front duration τ_f. By the current breaking moment τ_-, it is necessary that the resistance of the thyristor be a lot less than the resistance of the load ($R_H >> R_-$). When $R_H \approx R_-$, after the breaking of the DSRD current, the load current, due to the continuing modulation of the thyristor resistance, will slowly "drag out" until the final value $I_{mH} = U_0/(2R_H)$. From the given, condition $I_{mH} << I_m$ follows; that is, the current broken by the DSRD should significantly exceed the current switched into the load.

This method makes it possible, using a DSRD, to shape short (nanosecond) changes in the voltage or current with a "shelf" which is unlimited (as is the case for the normal switching on of a thyristor) with respect to duration of DSRD pumping. In the case of switching a load in parallel to the DDRV, the pumping circuit turns out to be parallel to the load, and sooner or later the transfer of the current between the load and the pumping current begins, which distorts the "shelf" of the pulse shaped by the DDRV.

1.2.2 Subnano and picosecond devices

Devices with delayed ionisation (Silicon avalanche shapers - SAS) are closing switches. As it was shown in previous chapter 1.1 to switch on SAS, constant voltage bias and fast rising voltage are needed. Fast rising voltage may be generated by one of the DSRD based circuits, described above. SAS basic circuits are shown at Fig. 1.9. Only part of DSRD based quasi -symmetrical circuit is shown. DC bias is provided by use of Cb, Lb,Rb parts. SAS may be connected in a gap in transmission line (Fig. 1.9a). In this case SAS only decreases ("eats out") pulse front, propagating through the line. If peaking capacitor Cs is connected, the output pulse amplitude may be increased up to two times. During the time of charging of Cs, considerable part of energy may be reflected from Cs and lost, that leads to poor efficiency of the circuit.

Space Charge region of a diode can store charge and DSRD may be used as a peaking capacitor (Fig. 1.9b) with no losses of energy due to reflection. Additional capacitor Cs may be connected in parallel to DSRD. All main parts (DSRD, Cb, SAS) may be assembled as one



SAS basic circuits

Fig 1.9

unit with minimal parasite parameters. Output voltage can not exceed the maximum voltage DSRD can hold.

Some modification of the scheme is shown at Fig. 1.9c. Storage (peaking) capacitor Cs is charged via inductor Lc. Cs charging voltage may be twice as high as on DSRD, but due to nonlinearity of capacitance of DSRD, the considerable part of energy stored in DSRD capacitance is lost (can not be transferred to Cs).

2 Fundamental limitations for semiconductor switches

2.1 A general approach to the problem of limitations

There are limits to performance of semiconductor switches. It is obvious that some limitations, for example those connected with the localisation of the current near the gate electrode, bear a special character which is determined by the type of devices, and can be lifted when changing for other types of the devices . For the given example, this can be done by means of the complete elimination of the gate electrode and use of two - electrode devices. At the very same time, fundamental limitations exist which are general enough in character, and which are determined by the parameters of the semiconductor itself, and also by the main mechanism for changing its conductivity, and are not related to the design of the specific devices.

An analysis of the fundamental limitations, which is the theme of this chapter, makes it possible to understand and evaluate from a single point of the view the degree of approximation of the real devices to the ideal; that is, to evaluate the main possibility of improving the parameters, and in some cases of finding the means of making this improvement as well. It is natural, that during the analysis we will always examine the case which is the most favourable for obtaining the maximum switching speed, and all evaluations will be done with funds "from above".

It is possible to represent the process of closing of switch in the following manner. As a rule, in the device, using the *p-n* junction biased in the blocking direction, an region is created in which there are no free current carriers. This region blocks the entire voltage applied from the external voltage source, and the current does not pass through the device. When filling the region with carriers, a current arises, which lead to the transferring of the power from the source to the load connected in series with the device. This process of filling with free carriers will hereafter be called modulation, while the filled region will be called the

modulation region (MR). It is possible to fill the MR with carriers by means of their injection from outside across the MR boundary, and also by means of the generation of carriers directly in MR using impact ionisation or external ionising radiation.

In agreement with the purpose given above, we will examine the limitations which are due to causes which are common for any device, regardless of its design. These causes are as follows: redistribution of the space charge due to the accumulation of carriers in the MR; electrodynamic phenomenon of the skinning of the current; the heating - up of the switch material and the mechanical stresses arising in this case.

We will deliberately exclude the slow diffusion processes of the transfer of carriers (the diffusivity is less than 10^2 cm^2/s), recombination, and also thermal processes (the thermal diffusivity is approximately 1 cm^2/s) from our consideration. Since large density currents are of interest to us, we will also disregard the effect of the space charge of ionised impurities. Let us look into the case which is of the most interest for practical work, in which the MR has flat parallel boundaries, and the current lines are parallel and directed along the x axis. We will assume that the electrons and the holes are distinguished only by the charge sign. This assumption will not disrupt the generality of the obtained conclusions.

2.2 The redistribution of the space charge

The process of filling the MR with current carriers is described by the following system of equations:

$$q\frac{\partial p}{\partial t} = -\frac{\partial jp}{\partial x} + qG \tag{2.1}$$

$$q\frac{\partial n}{\partial n} = \frac{\partial j_p}{\partial x} + qG \tag{2.2}$$

$$\frac{\partial E}{\partial x} = \frac{q(p-n\pm N_{d,n})}{\varepsilon} = \frac{\rho}{\varepsilon} \pm \frac{qN_{d,a}}{\varepsilon} \tag{2.3}$$

where G is the carrier generation function; ρ is the density of the space charge of carriers; $N_{d,a}$ is the concentration of the donor or acceptor impurities; ε is the dielectric permeability.

From equations (2.1) and (2.2) it follows that

$$qv\frac{\partial(p+n)}{\partial t} = -v\frac{\partial(v\rho)}{\partial x} + 2qvG \tag{2.4}$$

Here the condition $\gg N_{d,a}$ is used. Taking into consideration that the velocity of the carrier transfer v can not be greater than the saturated drift velocity v_s, we get

$$\frac{\partial j}{\partial t} = j' \le j'_m = -v_s^2\frac{\partial p}{\partial x} + 2qv_sG \tag{2.5}$$

where $j = qv(p + n)$; the stroke in the index designates the derivative with respect to time.

Let us examine a purely injection mechanism for filling the MR with current carriers ($G = 0$). For switching purposes, the conduction current is important, and the best case is that in which the entire current is the conduction current; that is, there is no displacement current, or to be more specific it can be disregarded. Then, as follows from equations (2.1) and (2.2), the density of the electron and hole current changes with the co-ordinate due to the accumulation of carriers in the MR (Fig. 2.1). In this case, only one point exists, in which the density of the currents of the electrons and holes evens out ($j_p = j_n$), while the density of the space charge of the mobile carriers equals zero ($\rho = 0$). We will accept this point as the beginning of the co-ordinates: x= 0. It is obvious that $\rho > 0$ when $x < 0$, and $x > 0$, $\rho < 0$ will be the case. As follows from equation (2.3), at point x = 0, the intensity of the electrical field reaches the maximum value E_m, dropping monotonously toward the MR boundaries to value E_- when $x = x_-$ and E_+ when $x = x_+$. When integrating equation (2.5) twice within the limits 0 -x and $x - x_+$ we find

$$j'_m \frac{x_+^2 - x^2}{2} = \varepsilon v_s^2 (E - E_+) \tag{2.6}$$

When substituting the designation $U_+ = \int_0^{x_+} E dx$ and integrating equation (2.6) within the limits 0 - χ_+, we finally get

$$j'_m = \frac{3\varepsilon v_s^2}{x_+^3}(U_+ - E_+ x_+) \tag{2.7}$$

By the very same means, for the interval x_ - x we find expressions which differ from formulas (2.6) and (2.7) only by the change in the "+" index for "-". A complete voltage drop on the entire MR with a thickness W=x+ +x_ equals $U = U_+ + U_-$.

Thus, the rate of increase in the current density will be limited by the greatest value for the x_+ or x_- expression. The maximum j' is reached in the symmetrical case, when $x_+ = x_- =$ W/2. From this condition, we find

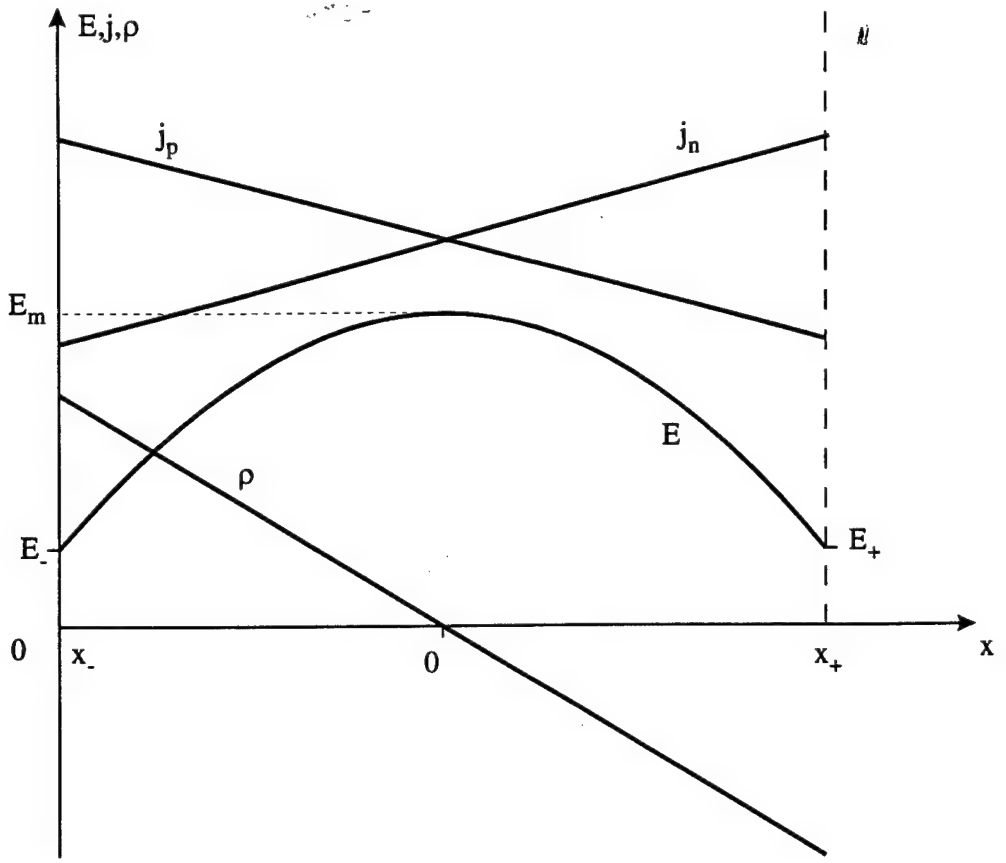
$$j'_m = \frac{12\varepsilon v_s^2 (U - \frac{E_+ W}{2})}{W^3} = \frac{8\varepsilon v_s^2 (E_m - E_+)}{W^2} \tag{2.8}$$

In the sharply asymmetrical case, when there is a large difference in the currents (for example, one-sided electron injection, $j_p << j_n$), the maximum field intensity turns out to be on one of the boundaries, and precisely opposite that boundary through which the carriers are injected. As a result of a calculation which is similar to the one done above, we get

$$j'_m = \frac{3\varepsilon v_s^2 (U - E_+ W)}{W^3} \tag{2.9}$$

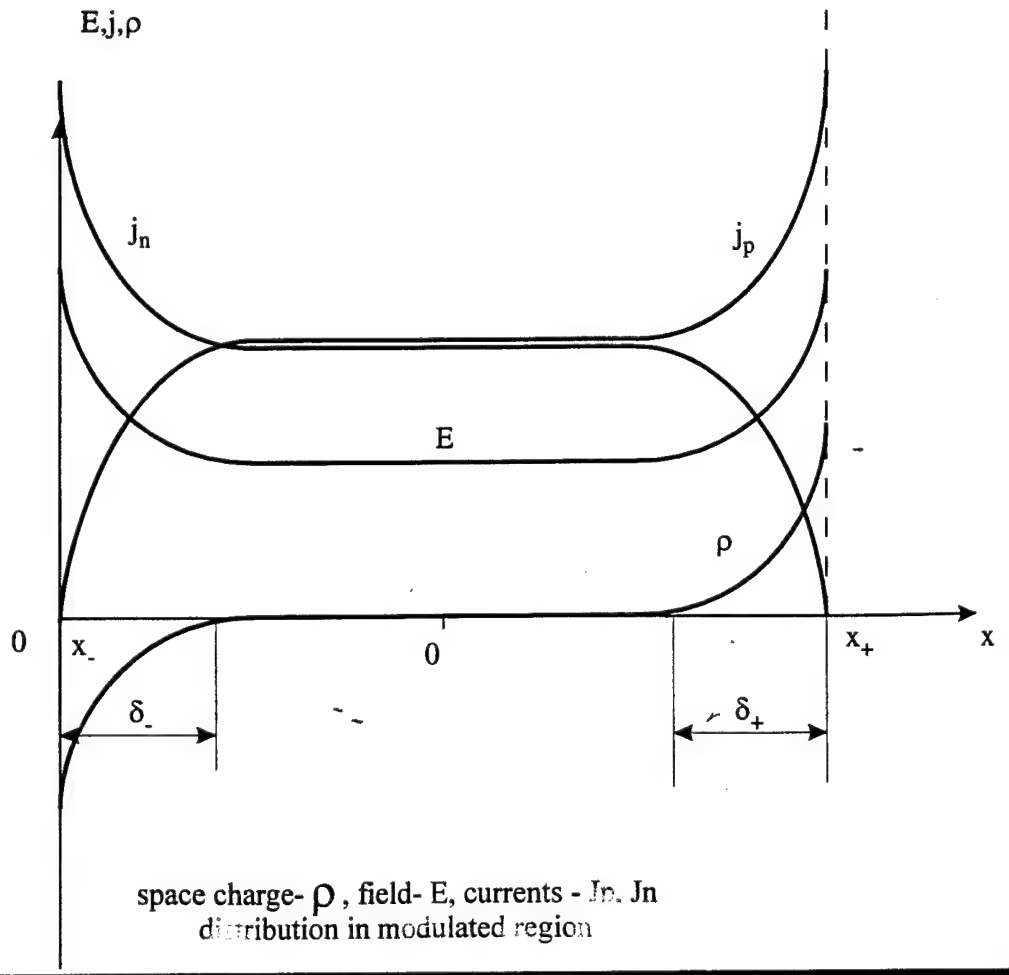
a

carriers injection



b

carriers generation



space charge- ρ , field- E, currents - Jp, Jn
distribution in modulated region

Thus, the maximum j' value in the asymmetrical case turns out to be four times lower than in the symmetrical case. From formulas (2.6) and (2.7), there also follows

$$E_m = \frac{3U}{2W} - \frac{E_+}{2}$$

The obtained expression for j' corresponds to the following physical pattern.

The increase of the current through the MR when there is a limited speed for the transfer of carriers is connected with an increase in their concentration. The concentration of the carriers, just as with the current, is distributed not uniformly along the x axis, while the maximum concentration will be at that MR boundary through which these carriers are injected.

The space charge which appears causes the redistribution of the electrical field. In this case the intensity of the field in the near - boundary regions decreases, and, in agreement with the last equation, the maximum field intensity in the central region increases. Zero field intensity at the MR boundaries corresponds to the threshold j'_m value. In this case, the subsequent increase in the current density is limited by the slow diffusion transfer of the carriers through the weak field areas which had appeared. Thus, in order to obtain the maximum rate of growth for the current, it is necessary to create an unlimited carrier source at the boundaries of the MR during switching.

It should be noted, that, when G=0, in equations (2.1) - (2.4) the sign of the term $\frac{\partial}{\partial t}$ at the left side will be opposite if the direction of currents are changed. That means: all results of the above made considerations of the plasma injection case may be applied to the case of plasma dispersal, i.e. current breaking devices.

In the case of the ionised generation of carriers, they appear in the MR in pairs. A space charge arises due to the separation of these pairs in the electrical field (Fig. 2.1b). It is obvious that the dimension which arose during time interval t_δ for the space charge region is evaluated by means of value $\delta \approx v_s t_\delta$, while the density of the space charge is $\rho \approx j/v_s$. The voltage drop U_δ in area δ is evaluated in the following manner:

$$U_\delta \approx \int\int_\delta \rho d^2x = \frac{j\delta^2}{2\varepsilon v_s} \qquad (2.10)$$

In the case of generation which is homogēnous with respect to the MR volume, and in the case of the absence of injection through the x_+ and x_- boundaries, a space charge will appear by each of the boundaries, and the distribution of the field will change. Because of this, two regions of increased field intensity will arise near the boundaries, and between them

there will appears region of lowered intensity. It is obvious that when fulfilling the condition $U = 2\ U_\delta$, the intensity in the area of the weak field will decrease to zero, which will also limit a subsequent increase in the current. Thus, from expression (2.10) we get the condition

$$j < j_m = \frac{\varepsilon v_s U}{\delta^2} \tag{2.11}$$

for which it is possible to disregard the effect of the space charge on the increase in the current.

As in the case of the injection mechanism for filling the MR, in the case of the ionisation mechanism, the limitation of the rate of growth of the current is determined by condition (2.5). However, which is easy to understand from Fig. 2.1 a, b the first component in this expression, which is connected with the effect of the space charge, has a negative sign, and can compensates for the second component, which describes the generation of carriers. There is a $d\rho/dx$ value for which it is possible the decreasing of the velocity j' to zero and the establishment of a stationary state with the maximum current density, limited by condition (2.11). As follows from expression (2.5), the characteristic width of the space charge region

$$\delta = \frac{j}{2qG} \tag{2.12}$$

corresponds to this state, taking into consideration the evaluation made above for ρ.

Taking expression (2.12) into consideration, we get from condition (211).

$$j < j_m = (4\varepsilon v_s U q^2 G^2)^{1/3} \tag{2.13}$$ As

is stated above, the first component in expression (2.5) compensates for the second component; therefore the j' value is the maximum at the beginning of the MR modulation process and is limited by the condition

$$j' < 2qv_s G \tag{2.14}$$

In the case of the generation of carriers by means of ionising radiation (photons, electrons, etc.), the generation function has the following appearance:

$$G = \frac{\phi\gamma}{\lambda}$$

where ϕ is the density of the flow of ionising particles; γ is the quantum output; λ is the depth of the penetration.

The maximum current density and the rate of its growth are determined by the substitution of the given expression into formulas (2.13) and (2.14):

$$j < j_m = (4\varepsilon v_s U q^2 \gamma^2 \phi^2/\lambda^2)^{1/3}$$

$$j' < j'_m = 2qv_s\gamma\phi/\lambda \tag{2.15}$$

For ionisation, the carrier generation function has the following appearance:

$$G = j\alpha(E)/q \qquad (2.16)$$

where $\alpha(E)$ is the impact ionisation coefficient.

The impact ionisation generation depends on the intensity of the electrical field, when there is a change in the distribution of the field (under the effect of a space charge), the homogeneity of the generation is also disrupted, and in particular generation stops in the area of zero intensity. Nevertheless, the discussions given above make it possible to easily obtain an evaluation from above for the current density and the rate of its growth. When substituting expressions (2.16) and (2.13) into formula (2.14), we get

$$j_m \approx 4\varepsilon v_s U\alpha^2;\ j'_m \approx 8\varepsilon v_s^2 U\alpha^3 \qquad (2.17)$$

As is well known, in a semiconductor, the increase in the impact ionisation coefficient is limited by its saturated value α_∞, which provides a natural threshold for expressions (2.17).

2.3 . The electrodynamic processes

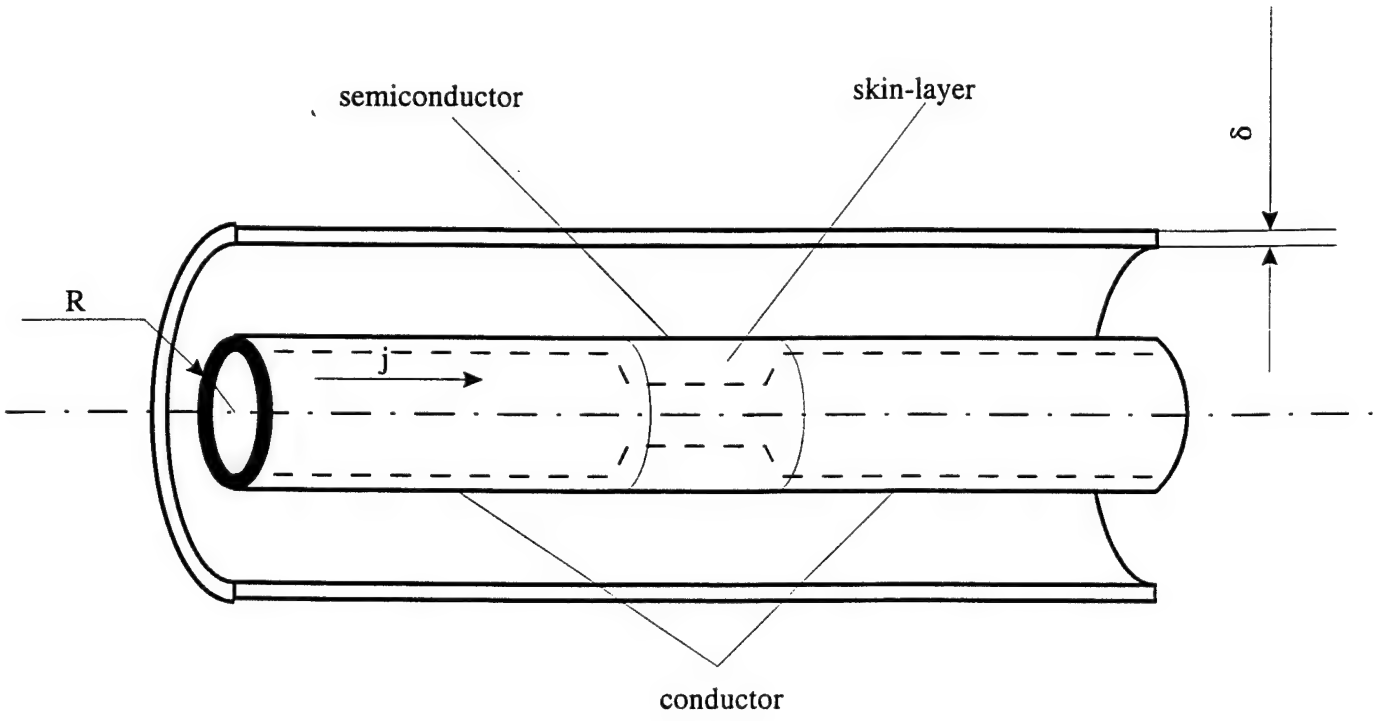
As follows from the Maxwell equations

$$rot\ \mathbf{H} = \qquad (2.18)$$

$$rot\ \mathbf{E} = -\mu_H \frac{\partial \mathbf{H}}{\partial t}\ , \qquad (2.19)$$

where H is the intensity of the magnetic field, and μ_H is the magnetic permeability, the change in the current flowing through the MR causes a change in the magnetic field which, in its turn, induces the electrical field; and this very electrical field can change the initial distribution of the current carriers. As is well known, in the transmission lines (waveguides), this process leads to the displacement of the current and the electrical field at the periphery of the conductor (skin effect). In a switch, before beginning the switching process (transition from the nonconducting state to the conducting state) there is no current through the MR, while the electrical field occupies the entire MR volume. At the end of the process, the switch is a conductor in which the skin effect is observed. It is obvious that in the switching process, the displacement of the field to the periphery occurs. Since the skin layer in the metal waveguide is thinner than in the switch, then at the boundary of the switch and the waveguide the current will be more strongly "clamped" to the surface (Fig. 2.2).

When designing systems with a high rate of change in the currents and voltages, one strives to come down to the minimum parasitic concentrated inductance and capacities. For this purpose, the geometry of the switch should be brought into agreement with the geometry of the waveguide; that is, - the structure of the electromagnetic field of the waveguide at the



skinning in fast switching

Fig. 2.2

site of the switch should not be disrupted. Naturally, this condition can be fulfilled for only one of the states of the switch, for example when switching the device on in the transmission line gap, which is for the conducting state. In the case of a coaxial waveguide, the switch should be cylindrical in shape, and in the case of a flat waveguide, it should have the shape of a belt, etc.

Let us examine MRs which are cylindrical in shape with radius R (Fig. 2.3), with current lines directed along axis x, and also which are in the shape of a belt with length L (Fig. 2.4). Disregarding the transitional region at the MR - conductor boundary yields an overstated evaluation of the thickness of the skin layer, and, consequently, of the threshold of the switching speed as well.

From expressions (218) and (2.19) it is easy to get

$$\text{rot rot } \mathbf{E}=\mu_H \frac{\partial \mathbf{j}}{\partial t} \qquad (2.20)$$

As was shown above, the effect of the space charge on the modulation process is expressed in the redistribution of the electrical charge along the x co-ordinate (along the current lines). It is obvious that, for this process, the characteristic longitudinal dimension χ is always less than the size of MR in the x direction; that is $\chi<w$. At the very same time, the characteristic transverse dimension of MR, that is the dimension in the direction perpendicular to the current lines (R and L in Fig. 2.3 and 2.4), is a great deal more than that in the longitudinal direction: for example $w \approx 10^{-2} cm, R \approx 1 cm$, that is $R >> W$. Therefore, the transversal components of the field intensity E_r is smaller than longitudinal (E_x) ($E_r << E_\chi$), that is, the lines of the field are practically parallel to the x axis. In this case, when disregarding the transversal components E_r and when assuming cylindrical symmetry ($\partial/\partial\varphi = 0$), expression (2.20) can be simplified in the following form:

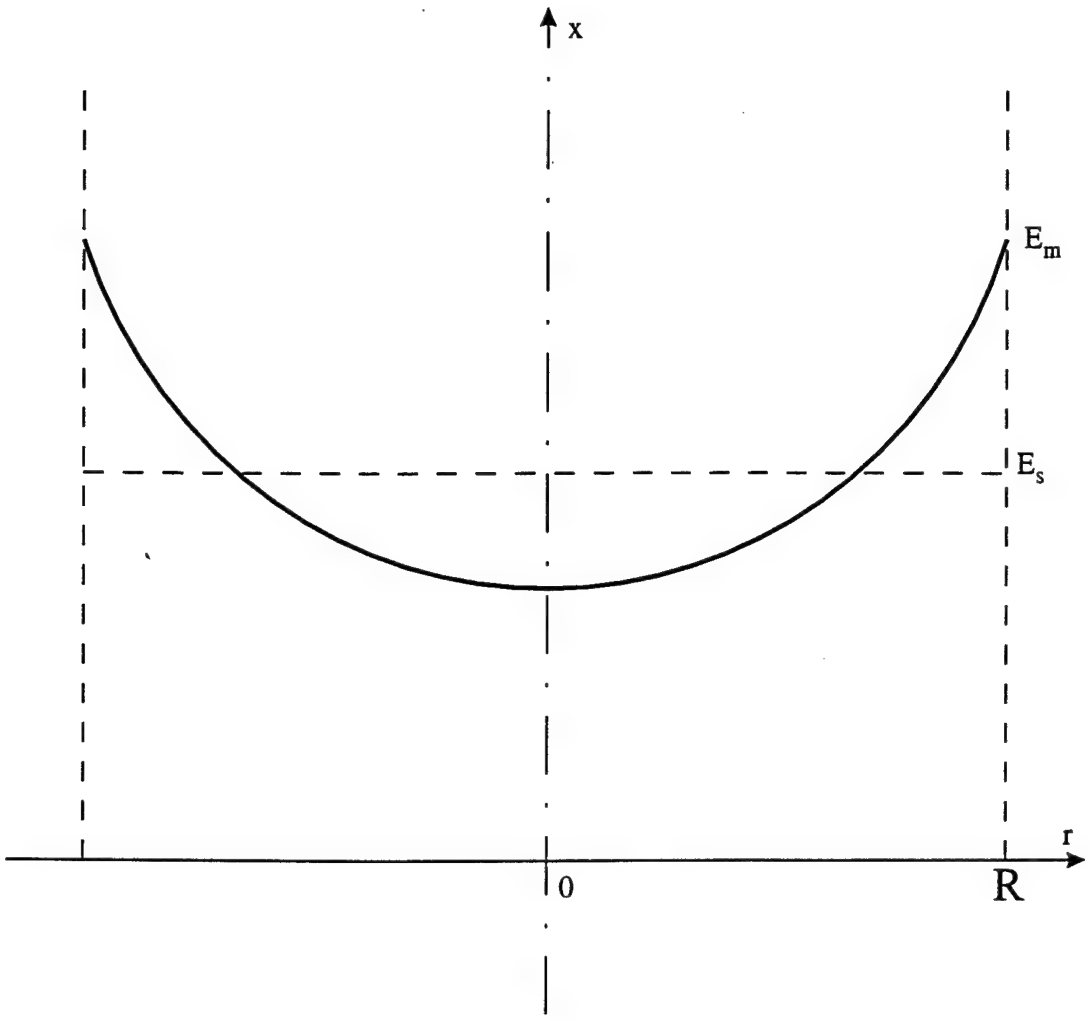
$$\mu_H \frac{\partial i}{\partial t} = \frac{1}{r}\frac{\partial}{r}\left(r\frac{\partial E}{\partial r}\right) \text{ when } \mathrm{E} \equiv E_x, \mathrm{j} \equiv j \qquad (2.21)$$

As will be shown below, in the case of large j' values, a strong non - uniformity of the field is possible in the radial direction with the characteristic δ size (skin layer), which is small as compared with $R(\delta << R)$. Nevertheless, it should be expected that equation (2.21) will continue to make sense when $\delta >> \chi$, that is, when there are large j' values, the longitudinal dimension χ can all the same remain smaller than the transversal dimension δ.

If the rise rate j' is given, then equation (2.21) can be easily integrated as follows:

$$E = E_m - \frac{\mu_H}{4} j'(R^2 - r^2) \qquad (2.22)$$

From here it follows that when the rate j' is given and $\partial j/\partial r = 0$, the intensity of the field drops from the maximum value at the edge of the MR ($r = R$) to the minimum in the centre. It is obvious there exist such critical values for the rates of growth of the current density and the complete current



skinning during modulation of conductivity in semiconductor

Fig. 2.3

$$j'_k = \frac{4E_m}{\mu_H R_0^2} \; ; \; I'_k = j'_k \pi R_0^2 = 4\pi E_m / \mu_H \tag{2.23}$$

when the field intensity in the centre, and consequently the current become equal to zero. That is, the accepted condition for the constancy of the rate j' along the radius will in no way be fulfilled when $j_k < j'$, and, in this manner, displacement of the current to the periphery begins, which can be interpreted as a skin effect for non-linear conductivity ($v = v_s \neq \mu_n E$).

An analogous analysis for an MR which is in the shape of a belt (Fig. 2.4) with two dimensions which differ sharply (L >>l) yields the following instead of expressions (2.23):

$$j'_k = \frac{4E_m}{\mu_H l^2}; \; I'_k = \frac{4E_m L}{\mu_H l} \tag{2.24}$$

It should be noted that the j'_k values determined by expressions (2.23) and (2.24) are not the threshold values for j'. When fulfilling the condition $j' > j'_k$ homogenous distribution of the current density along the MR area becomes impossible, and the current is displaced to the boundary region with characteristic dimension δ_μ, which is less than that of the transverse dimension of the MR. It is obvious from equation (2.21) that in the general case,

$$\delta_\mu \approx \sqrt{\frac{E}{\mu_H j'}}, \tag{2.25}$$

where E is the characteristic value of the field intensity, which with respect to its order is close to E_m.

It is easy to show that expression (2.25) is also valid for belt-shaped MRs.

Solution (2.22) to equation (2.21) is obtained under the condition of the independence of the j' value on the field intensity, which is correct only in the case of ionisation by means of external radiation, and in the case of a constant carrier drift velocity. Although for other modulation mechanisms the speed j' depends on the field intensity and equations (2.22-2.24) stop being fulfilled exactly the evaluation of (2.25) remains correct. The δ_μ expressions for different modulation mechanisms are easily obtained by the substitution of the j' expressions from paragraph 2.1 into formula (2.25), which will ;be done after we examine the skin effect and the heating of the material together.

2.4 The heating of the material

The passage of the conductivity current through the MR is accompanied by the dissipation of thermal energy Q_t and by an increase in the temperature for value $\Delta T = Q_\tau / c_\tau$,

$$Q_\tau = \int_0^t jE dt = \int_0^j \frac{jE}{dj/dt} dj. \tag{2.26}$$

where c_τ is the specific thermal capacity:

The maximum value for the energy density, which can be dispersed in the switch's material, is limited by the melting of the material or by its breakdown due to the mechanical stresses which arise from the thermal expansion.

In the first case, the Q_τ value is determined by the sum of the energy going for heating to the melting point, and the specific melting heat.

Let us examine the second factor in more detail in the example of a belt-shaped MR (see Fig. 2.4). Obviously, the smallest mechanical stresses arise during the uniform heating of an MR with mechanically free boundaries. In the stationary state, under these conditions, there are no mechanical stresses. However, in the dynamic state, the picture changes. Increasing the temperature by dT will cause expansion along the γ axis by value $d\gamma = a_\tau \gamma dT$, where a_τ is the thermal expansion coefficient. At point γ, pressure $P_\tau = m(d^2\gamma/dt^2)$ arises, where $m = g(L-\gamma)$ is the inert mass of a unit of area and g is the density of the material. From what has been stated it follows that:

$$P_\tau = a_\tau g y (L-y)\frac{d^2T}{dt^2}$$

The maximum pressure is found in the centre and, taking into consideration expression (2.26), equals

$$P_m = \frac{a_\tau g L^2}{4c_\tau}\frac{d^2Q_\tau}{dt^2} \approx \frac{a_\tau g L^2 E_c}{4c_\tau} j' \quad , \tag{2.27}$$

where E_c is the characteristic field.

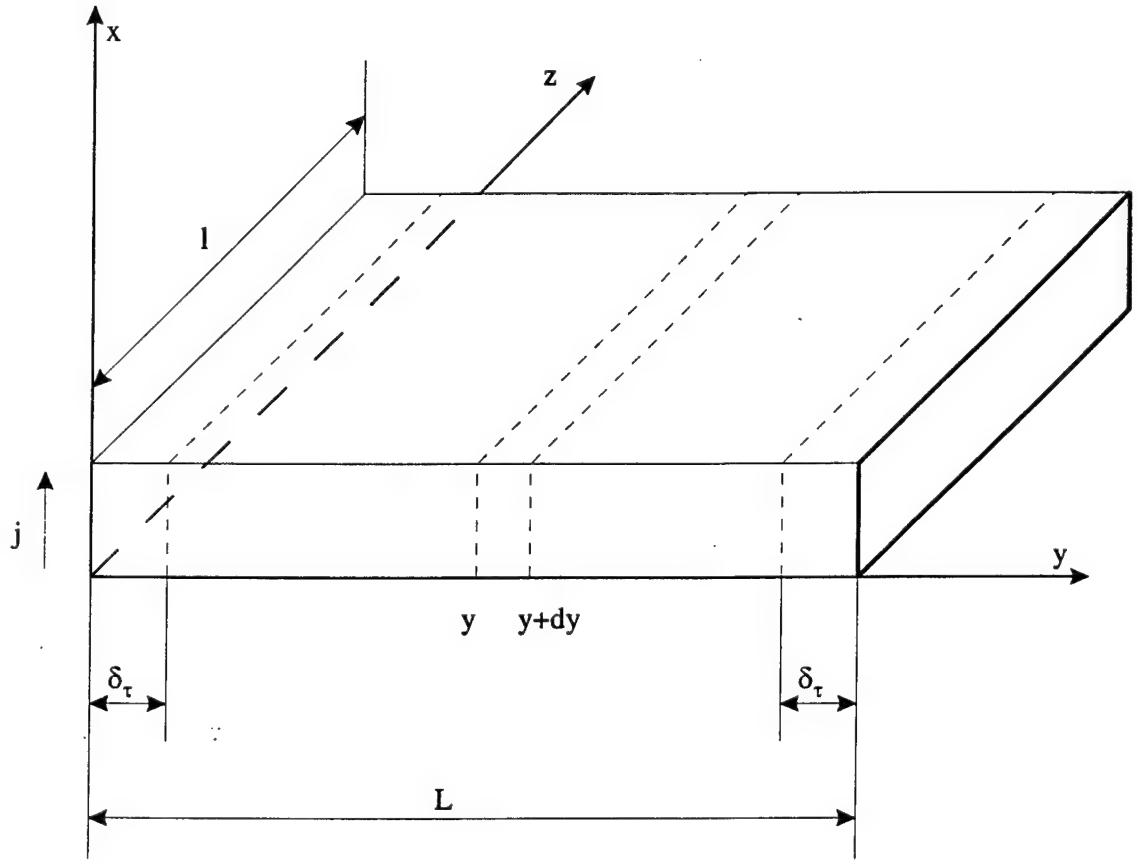
Thus, the maximum pressure is proportional to the rate of rise of the current density and expression (2.27) can be seen as a limit for j', after first ruling out E using expression (2.8) or (2.17). However, this limitation is not absolute.

It is well known that the maximum speed for the transmission of the pressure is the velocity of sound. For current rise time τ_j, which is small as compared to the time of the sound run $\tau_s = L/c_s$, (where c_s is the speed of sound) along the largest MR border, the shift of mechanically free boundaries along this border is determined only by the expansion of a layer adjoining it with thickness $\delta_\tau = \tau_j c_s$ (see Fig. 2.4). Therefore, when $\tau_j << \tau_s$ it is possible to consider the boundaries $\gamma = 0$ and $\gamma = L$ to be mechanically unmovable, and we get

$$P_\tau = \frac{a_\tau \Delta T}{\chi} = \frac{a_\tau Q_\tau}{\chi c_\tau} \quad , \tag{2.28}$$

where χ is the compressibility of the material.

As follows from expression (2.28), when there is a small switching time, which is of the main interest to us, the maximum value for the density of the heat liberation $Q_{\tau m}$ is limited by the stress of the breakdown of the material P_p (of course, if melting does not occur



dynamical mechanical stress in semiconductor

Fig. 2.4

earlier). It is obvious that the obtained evaluation for $Q_{\tau m}$ keeps its force for cylindrical geometry as well, with the substitution of L for $2R$.

When substituting value j' according to formula (2.8) into expression (2.26), we get the threshold value for the current density for an injection mechanism:

$$j_m \approx \frac{4v_s}{w}\sqrt{\varepsilon Q_{\tau m}} \ . \qquad (2.29)$$

The limitation for the current density (2.12) containing the regime parameter- -voltage *U- -has* already been obtained for an ionisation mechanism. When excluding U from equations (2.12), (2.13), (2.14) and (2.26), we get the limitation for the current density taking into consideration the heat losses:

$$j_m \approx (v_s q G)^{1/2}(2\varepsilon Q_{\tau m})^{1/4} \ . \qquad (2.30)$$

For an impact ionisation mechanism, excluding G from expressions (2.30), (2.13) and (2.16), we find

$$j_m \approx 4\alpha v_s \sqrt{\varepsilon Q_{\tau m}} \ . \qquad (2.31)$$

From formulas (2.26) and (2.25), taking into consideration that $I_m \sim 2\pi\delta R_{jm}$, it follows that

$$I_m/2\pi R \approx 4\sqrt{Q_{\tau m}/\mu_H} \ . \qquad (2.32)$$

Limitation (2.32) for the total switching current passing to unit of length of the perimeter of the MR bears a universal character, and is determined only by the threshold density of the dissipated energy. Expression (2.32) has the following physical meaning: when there is a decrease in the switching time, the current is displaced to the periphery, which leads to an increase here in the density of the current and the density of the dispersed energy. In this case, the current density increases in such a way that, in agreement with expression (2.32), the total energy dissipated in the form of heat does not depend on the switching time. This energy also does not depend on the conductivity of the MR, since, for example, when increasing the conductivity, the current is concentrated to a great degree at the periphery. Since condition (2.32) implies the smallness of the thickness of the skin layer in comparison with the radius ($\delta_\mu << R$, then, after turning a ring with thickness δ_μ along which the current flows, it is possible to make the transition to a belt-shaped MR, and to substitute 2π for L in expression (2.32) for the evaluation.

In the switching process, the power dissipated in the load $P = R_H I_2$ and the speed of the switching of the power is

$$\frac{dP}{dt} = P' = 2R_l II' = 2I'(U_0 - U), \qquad (2.33)$$

where U_0 is the voltage of the power source, Rl is the load resistance.

Thus, the threshold value for the power can be found from condition (2.32), and the threshold values P' which are needed for a unit of area can be easily determined from formula (2.33) and the corresponding expressions for j'. Let us make note of the fact that in those conditions when $j' \approx U$ [see formulas (2.17) and (2.18)], the maximum P' is reached when $U \approx U_0/2$.

Table 1.1

Limiting process	Limited value	Threshold value for different mechanisms for filling the MR		
		double injection	external radiation ionization	impact ionization
Space charge redistribution	dj/dt	$12\varepsilon v_s^2 U/W^3$	$2qv_s\gamma\phi/\lambda$	$8\varepsilon v_s^2 a_\infty^3 U$
Heat dissipation	j	$v_s(\varepsilon Q_{\tau m})^{1/2}/W$	$2\varepsilon Q_{\tau m})^{1/4} \times (qv_s\gamma\phi\lambda^{-1})^{1/2}$	$a_\infty v_s(\varepsilon Q_{\tau m})^{1/2}$
Skin effect	I/L	$4(Q_{\tau m}/\mu_H)^{1/2}$		

The results obtained above are given in table 1.1 and can be summed up in the following manner: the process of accumulating carriers which leads to the appearance of a space charge and to the redistribution of the electrical field superimposes limitations on the maximum achievable j' values. The heating up connects the threshold values j' and j. The electrodynamic effect of the displacement of the field to the periphery connects the limitations for j' and j with the threshold value for the current I which is possible for a unit of length of the MR's perimeter.

2.5 The evaluation of the threshold values of the switching parameters

It is well known that the difference in the parameters which are included in the formulas for table 1.1 for all widely used monocrystalline semiconductors (Si, Ge, GaAs, etc.) is rather small, and is within the limits of an order of magnitude. Therefore, further numerical evaluations for silicon which have been done are close to the evaluations for other semiconductors.

The breaking stress for silicon $P_\tau \le 10^5 N/cm^2$, which yields $Q_{\tau m} \le 2 \bullet 10^3 J/cm^3$ is the value which approximates the energy which guaranties heating up to melting. Thus, the

maximum switched current for a unit of length $I/L \approx 10^6 A/cm$, and does not depend on the MR filling mechanism.

As follows from table 1.1, the limitations on j' and j are like the limitations for injection and impact ionisation mechanisms. In the latter case, the free path between ionisation acts 1/ plays the part of the thickness W. Thus, the effectiveness of the impact ionisation mechanism with respect to the injection mechanism is determined by the value wa, which for the high - voltage devices which interest us ($U > 1kV, w \approx 10^{-2}cm, a_\infty \approx 10^5 cm^{-1}$) equals approximately 10^3.

When equating the j'_k values [see formula (2.23)] and the corresponding j'_m values, it is possible to obtain conditions for which skinning begins:

for an injection mechanism

$$v_s/c \approx W/ \quad . \tag{2.34}$$

for an impact ionisation mechanism

$$v_s/c \approx 1/(aR \quad . \tag{2.35}$$

where $c = (\varepsilon\mu_H)^{-1/2} \approx 10^{10} cm/s$ is the speed of the electromagnetic wave in the semiconductor.

Since $v_s/c \approx 10^{-3}$, in the case of the injection mechanism skinning can play a substantial part only in structures with an extremely large $R/$ ratio, for example when $w \approx 10^{-2}cm, R \approx 10cm$. At the present time, diameter for the structures used in nanosecond equipment reaches as much as several centimetres, and when there is a further increase in R, the skin effect should become noticeable.

For an impact ionisation mechanism, if $a \approx 10^5 cm^{-1}$, skinning can have an effect when there is an $R > 10^{-2}cm$ value which is completely real for practical purposes.

In the case of an injection mechanism, the threshold current density which is limited by the heat dissipation equals $j_m \approx 10^5 A/cm^2$ for the accepted values $v_s \approx 10^7 cm/s, w \approx 10^{-2}cm$ and $j'_m \approx 10^{12} A/(cm^2 \bullet s)$ when $U \approx 10^3 V$. It should be noted that these j_m and j'_m values are relatively small. Thus, fast-response thyristors, the thickness of whose MR was $w \approx 1.5 \bullet 10^{-2}cm$ [5], were studied; when the load resistance $R_H = 10ohm$ at a moment in time which corresponds to the voltage at the device $U \approx 500V$, the following values are obtained: $I' = 5 \bullet 10^9 A/s, j' = 4 \bullet 10^{12} A/(cm^2 \bullet s), j_m \approx 8 \bullet 10^4 A/cm^2$; that is, in high-power switches, j' and j values which are close to the threshold values for an injection mechanism have already been realised, or they already exceed these (j').

For an impact ionisation mechanism, the corresponding j'_m values in $(wa_\infty)^3 \approx 10^9$ times, while j in $wa_\infty \approx 10^3$ times greater than for an injection mechanism; there $a_\infty \approx 10^5 cm^{-1}$ is the saturated value for the impact ionisation coefficient in silicon.

The previously given division of the mechanisms for filling MRs is conditional in character. In reality, their sequential replacement or simultaneous existence is possible. Thus, as a consequence of the redistribution of fields when there is an injection mechanism or ionisation by means of external radiation, an impact ionisation region can arise as the current increases.

The high j' and j values noted above, which were realised in thyristors, are connected with the transition from an injection mechanism to an impact ionisation switching-on mode.

The special role which impact ionisation plays or can play in switching processes is obvious from what has been stated. First of all, it frequently arises in fast, high-power switching processes; secondly, this is an extremely powerful source of carriers. As follows from formulas (2.16) and (2.17), the maximum value for the generation function $G_m \approx 10^{32} cm^{-3}/s$. For the sake of comparison, we point out that, during optical generation, such a G value is equivalent (for light with an absorption depth of $\lambda \approx 10^{-3} cm$) to the light flux $\phi \approx 10^{29}$ photons/(cm^2 s) or $\phi \approx 10^{10} W/cm^2$ when the photon energy equals approximately 1 eV.

In the analysis given above, we did not examine tunnel interband generation. In very strong fields ($E \approx 10^8 V/cm$), when the probability that a barrier will be passed by a valent electron is close to one, the rate of the tunnel ionisation can significantly (by many orders of magnitude) increase the G_m value given above. However, in sufficiently protracted intervals, the voltage drop on which exceeds the impact ionisation threshold, pure tunnel ionisation is unrealisable. As the intensity of the field increases, even to the point that $E \approx 10^6 V/cm$, and when the rate of the tunnel ionisation is still small, the impact ionisation coefficient reaches the saturated value a_∞. Therefore, the very first carriers which came to be by means to the tunnel mechanism will cause a more intensive impact ionisation avalanche.

The examination given above of the limitations on the commutation parameters I, j, j' did not consider items which are so very important for practical purposes, such as the simplicity and effectiveness of the realisation, the possibility of the appearance of any type of instabilities which worsen the reliability and stability of the performance of devices, etc.. From this point of view, the prospects for using an impact ionisation mechanism connected with strong carrier heating are less definite as compared with those for the injection mechanism.

3 Properties and limitations of submicrosecond switches

As was shown in section 1.2 new step recovery diodes are used with primary closing switches, which must produce current pulses in submicrosecond range . The set of parameters of the switches strongly determines the potentialities of pulse forming circuits. After one cycle of pumping and current break a DSRD is ready for the next cycle and the period of the pulse repetition limited by DSRD is hundred or several hundreds nanoseconds only. That leads to the maximum possible repetition rate in megahertz range. In actual circuits maximum repetition rate is limited by primary switches: for power thyristors - dozens kilohertz only. Power Field-effect transistors (FET) could provide far higher repetition rate as high as megahertz. The other important parameters for the primary switches are turn-on time, resistance in turn-on state, dI/d capability.

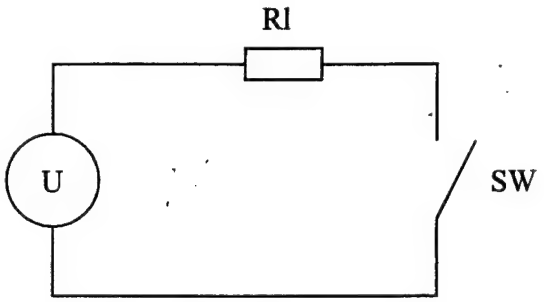
The border between transient turn-on process and steady (static) turn-on state is rather fuzzy. Typical for every kind of switch, turn-on process of switching of constant voltage into R is shown on Fig. 3.1. There is initial fast voltage drop from off-state voltage, then the rate of voltage drop (dU/dt) sharply decreases by order of magnitude. The length of this slow "tail" may be tens times more than the first fast part. Gradually "tail" comes to steady state voltage drop U_{on} . "Tail" voltage drop may be tens times higher than steady-state drop.

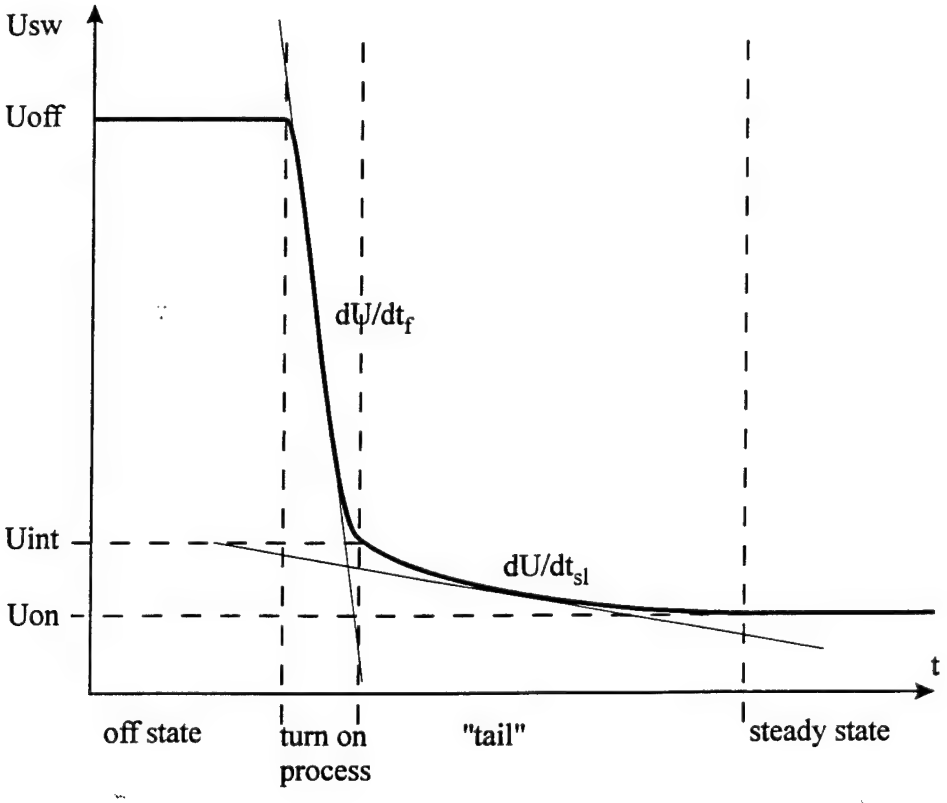
As a rule (after turn on) power high voltage devices don't reach steady state during relatively short time (100-300 ns) while DSRD being in high conducting state ($\tau_- + \tau_+$. Therefore well known for the devices parameter "Static On-Resistance" is not of much value for losses of energy in the switch evaluation.

In all circuits discussed in section 1.2 the primary switches have to provide needed current rise rate ($I' = dI/dt$. For LC circuits with maximum current I_m and oscillating frequency f $\{I = I_m \bullet \sin(2\pi ft)$ intrinsic value I'_i is $I'_i = 2\pi f I_m = \pi I_m / T_h$, where T_h is the halfperiod time.

It may be shown that high Q-factor is possible only when dI/d capability of the primary switch I'_s is much larger than LC intrinsic value ($I'_s > I'_i$. When $I'_s < I'_i$ current oscillation in LC - *circuit* is impossible, (Q factor is less than one), most of energy stored in capacitor dissipates in the switch.

In the next sections we consider the properties of semiconductor power switches transistors, thyristors and dynistors for use as a primary switches. We will not discuss the triggering systems that drive switches.





turn on process in semiconductor closing switch

Fig.3.1

3.1 Transistors

3.1.1Bipolar transistors

The physics of bipolar transistor's operation is well known. In turn off state applied voltage is blocked by collector *p-n* junction (Fig. 3.2). The width of space charge region W_{SCR} is

$$w_{scr} = \sqrt{\frac{2\varepsilon U}{qN_d}} . \tag{3.1}$$

In turn-on process, electrons are injected into SCR via p-base layer. The collector and load current increases. Due to voltage drop on the load the voltage drop on p^+nn^+ layers of transistor decreases when the transistor current increases.

There are two main factors limiting current density rise rate (*dj/dt*). One is due to diffusion of electrons via p-layer. The collector current density (j_c) rise in case of ideal efficiency of emitter $\gamma = 1$ may be approximated by

$$j_c = j_b \frac{\tau_n}{\tau_d}[1 - \exp(-t/\tau_n)]. \tag{3.2}$$

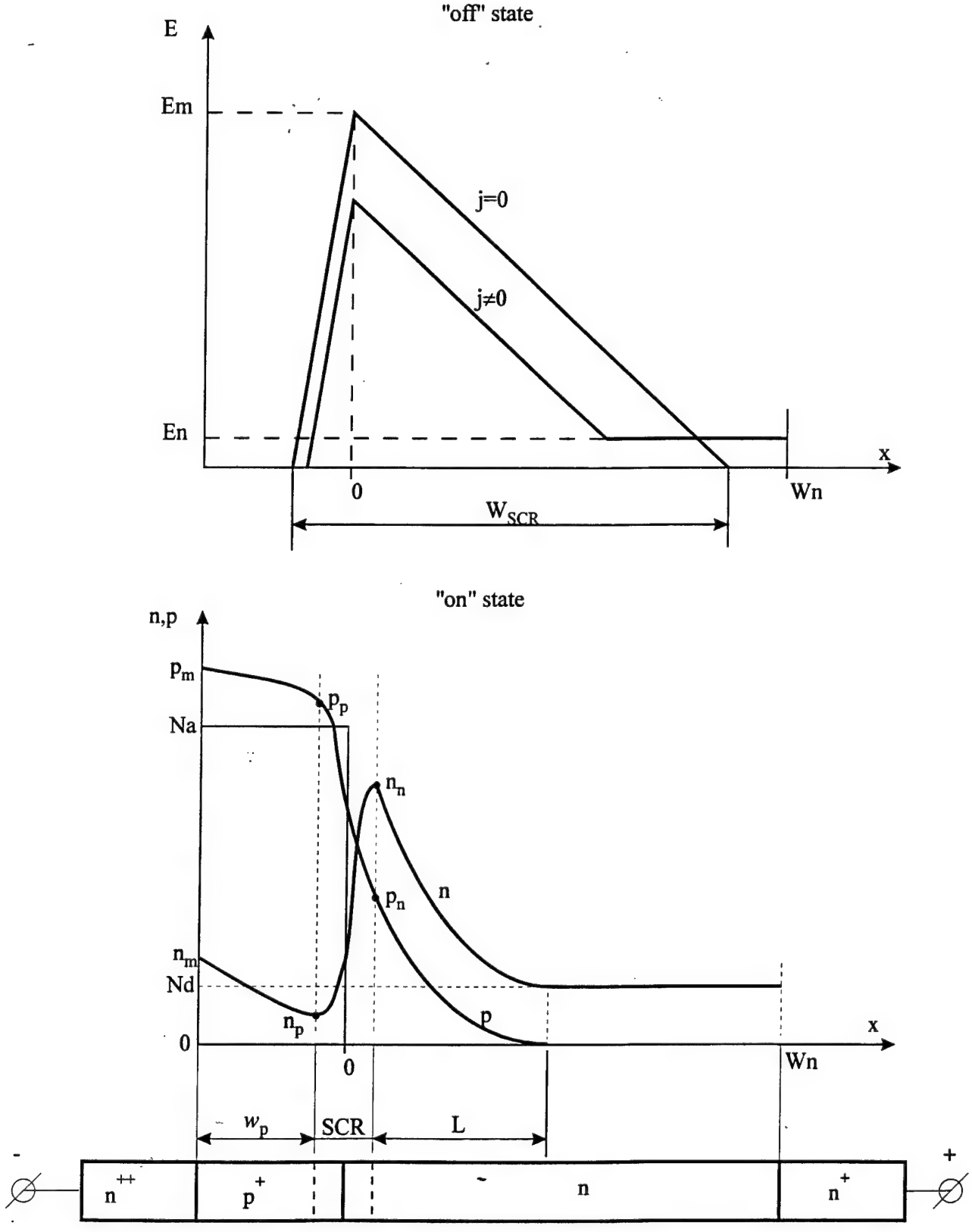
where τ_n - electrons time of life in p+ - layer, $\tau = \frac{W_p^2}{2D_n}$ - time of diffusion of electrons through p+ - layer, D_n - diffusivity, j_b - gating base current.

From (3.2) for maximum value of dj/dt_m we have:

$$dj_m/dt_m = j_b/\tau_b . \tag{3.3}$$

High voltage transistors, as a rule, have p -layer width near 10 μm and τ_d near 20 ns. Even for large effective collector area $S \approx 1cm^2$ and base current $\approx 10A$ we have $I'_m = dj_c/dt_m \approx 0,5 \bullet 10^9 A/s$. This value should be compared, for example, with $I'_i = 2 \bullet 10^9 A/$ *LC* circuit, needed to generate 100 A pulse ($T_h \approx 50ns$. It is evident that more than 10 transistors must be used in this example to get good *Q*-factor. It is possible to increase I'_m capability by decrease of W_p and τ_d. For $W_p \sim 1\mu m$, τ_d is less than 1 ns. For such short times of diffusion (less than time of flight of electrons across space charge region) I'_m capability is limited by space charge redistribution for the case of one side injection, considered in section 2 .(2.9) gives $j'_m \le 10^{11} A/cm^2 \bullet s$).

In accordance with (3.1) $W_{SCR} \approx 10^{-2} cm, N_d \approx 10^{14 cm^{-3}}, \tau_f = W_{SCR}/V_S \approx 1ns$ for transistor blocking more than 1 kV. Such transistor should have > 100 times better I'_m capability, but to our knowledge they are not produced. It should be noted, that so thin base transistors will be very close and similar to high voltage vertical field effect transistors, where carriers transfer through thin low field region does not limit the current rise.



field and carriers distribution in bipolar transistor

For given collector current density during turn on process transistor voltage drop consists of two part: space charge region voltage drop (3.1) and neutral region voltage drop (U_n) (Fig. 3.2)

$$U_n = \frac{j_c \bullet W_{nut}}{q \bullet \mu_n \bullet N_d}, \tag{3.4}$$

where W_{nut} is neutral region width. The space charge voltage drop may be determined as:

$$U_{SCR} = U - Rl \bullet S \bullet j_c - U_n, \tag{3.5}$$

where $Rl \bullet S \bullet j_c$ - load resistor voltage drop.

The expression (3.4) is valid when $j_c < j_s = qV_sN_d$. If $j_c \approx j_s$, space charge of electrons compensated space charge of donors in *n - layer*, SCR widens up to n^+ layer and neutral region disappears. If $j_c > j_s$ electric field gradient changes sign, field maximum is shifted to the *n+ - layer*, and field intensity at collector *p+n - junction* decreases. This field decrease slows down the electron drift velocity at p^+n-junction and decreases the rate of current rise (dI/dt . So, the current density should be less than $j_s = qV_sN_d$.

Nevertheless while $j_c < j_s$, it is possible, that in accordance with (3.5) U_{SCR} goes to zero or even changes signs. That means the appearance of diffusion region at *p-n* junction instead of high voltage *SCR* (Fig. 3.2 "on state"). The width of diffusion region (L) increases

$$L \approx \sqrt{2Dt,} \tag{3.6}$$

and the width of neutral region where voltage drop is determined by (3.4) decreases. The voltage drop at diffusion region is very small (less than 1 V).

The rate of width increase (3.6) is diffusion slow process. As follows from (3.6) it takes diffusion region ~ 2 microseconds to fill *n - layer* of ~ 100 micron width. It is the process that determines the slow "tail", mentioned above (Fig. 3.1).

Hence the main, well known advantage of bipolar transistors in low static on-state voltage drop over field effect transistors plays no part in their use as a primary high voltage switches. In low voltage high current transistors (for 100 V, $W_n \approx 10^{-3}$ cm and filling times ≈ 20 ns are possible) this advantage may play significant role.

For 1 kV transistor with $S = 1cm^2, W_n \approx 10^{-2}cm, N_d \approx 10^{14}cm^{-2}$ neutral region voltage drop as high as ≈ 50 V at 100 A current is possible. Associated dynamic resistance is approximately 0.5 Ohm, that may strongly limit Q- factor for LC circuits with impedance ($\rho_{LC} = \sqrt{L/C}$ less than 10 Ohm.

As was mentioned in section 1.2, total efficiency of use of switches as a primary switches may be defined as ratio of maximum currents during the first (I_1) and the second (I_2) halfperiod (T_h) of oscillation in LC circuit with the switch. It should be noted that the task to determine experimentally Q factor from the decay of oscillations is poor defined when Q factor is low, less than 5.

The devices, tested in LC circuit, must be shunted by a diode to bypass reverse current. The diode has to have low forward resistance not to increase losses and to diminish Q-factor. Such diodes were specially developed by us and are in considerable use with all kinds of primary semiconductor switches. Our tests of bipolar transistors in LC circuits have verified the considerations made above. For example test of 1 kV rated *npn* transistor KT854 (peak current $I_m = 18$ A, $U_{off} = 400$ V, driving base current $I_b = 3$ A, $T_h = 160$ ns, $W_p \approx 20\mu m$) has showed $I_1/I_2 = 1{,}5$ (poor Q factor $\approx 2 \div 3$). Increase of driving current improved Q factor, but not significantly, due to decrease of transistor gain at large currents.

Low voltage (< 100V) transistors KT970 have better Q factor $\approx 5 \div 7$ at peak current $I_m \approx 30A$. This Q factor allows the use of this transistor as primary switches.

3.1.2 Field effect transistors

High voltage vertical field effect transistor (FET) have much in common with bipolar transistor: space charge region, blocking applied voltage, injection of carriers into the SCR, that leads to current increase and so on. The main difference is the source and way by which the carriers injected into SCR. The time needed for electrons to get from source into SCR, even in high voltage FET, is far less than in bipolar transistors and is as small as several nanoseconds. This times are close to time of flight across SCR and considerations, made above for the case of small τ_d , may be applied to FET: (3.4), (3.5), (2.9) are valid.

When voltage drop at SCR decreases and U_{SCR} in accordance with (3.5) goes to zero no diffusion region appears contrary to bipolar transistors. Nevertheless electrons transfer across the low field region slows down and rate of current rise decreases. Minimal voltage drop is determined by drop at neutral region (3.4) not enriched by electron-hole plasma. FET's resistance in static turn-on state is higher than in bipolar transistor. As was shown in previous section this advantage plays no role in case of use of high voltage devices in LC circuits with $T_h \sim 10^{-7}$ sec . Due to faster transfer electrons from source to SCR, FET have better dI/d capability than bipolar transistors with usual thick *p - layer*. We could remember, that bipolar transistors with thin p - layer would have the same I' capability as FET and could provide the same and may be slightly better Q-factor due to lower static "on" resistance .

The tests of FETs in the same as above LC circuit (T_h = 160 ns) showed:

a) 1 kV rated IRFBG30

ρ = 50 Ohm, I_1/I_2 = 1,15, Q ≈ 5÷7, I_1 = 8A

ρ = 24 Ohm, I_1/I_2 = 1,5, Q≈ 2 ÷ 3 ,I_1 = 16

b) 500 V rated IRF840

ρ = 24 Ohm, $I_1/I_2 \approx 1,3$, $Q \approx 3 \div 4$, $I_1 = 16$

ρ = 18 Ohm, $I_1/I_2 = 1,5$, $Q \approx 2 \div 3$, $I_1 = 22$

The results show that decrease of LC impedance worsens Q-factor and peak current increase . The more voltage rated is the more static "on" resistance, and *n - layer* voltage drop as follows from (3.1), (3.4) (provided the same effective area S).

3.2 Four layers devices.

The most powerful semiconductor devices are devices based on four layers $n^+p^+np^+$-structures - thyristors (if they have third gating electrode) or dynistors (if they have only two electrodes). Although they have only one *p+ - layer* more than bipolar transistor, the physics of their operation is changed dramatically. For power high voltage thyristors no good grounding in theory exists. There are only several hypothesis explaining each specific feature of thyristor's behaviour. The distinctive property of thyristor is the double injection: electrons are injected from *n+- emitter*, holes are injected from opposite placed *p+- emitter*. Such two-sides injection may fill all volume of *n- layer* by high density electron-hole plasma during much shorter time than in case of transistor. The space charge redistribution limitation considered in section 2.1 (2.8) gives for dj/dt factor in symmetrical case (thyristor) 4 times more value, than in asymmetrical (transistor). In thyristor the direction of drift component of the injected carriers velocity coincides with diffusion component in *p+ -layer* (base). Due to this coincidence electron's time of flight across rather thick (> 10μm) *p+ - base* may be very short at high current density. In this case heavy dj/dt limitation (3.3) connected with *p+ - base* transfer time plays no role. As was mentioned above the limit defined by (2.8) $dj/dt \approx 10^{12} \frac{A}{cm^2 s}$ was achieved in thyristors. This value is many orders of magnitude more than achieved in bipolar or field effect transistors.

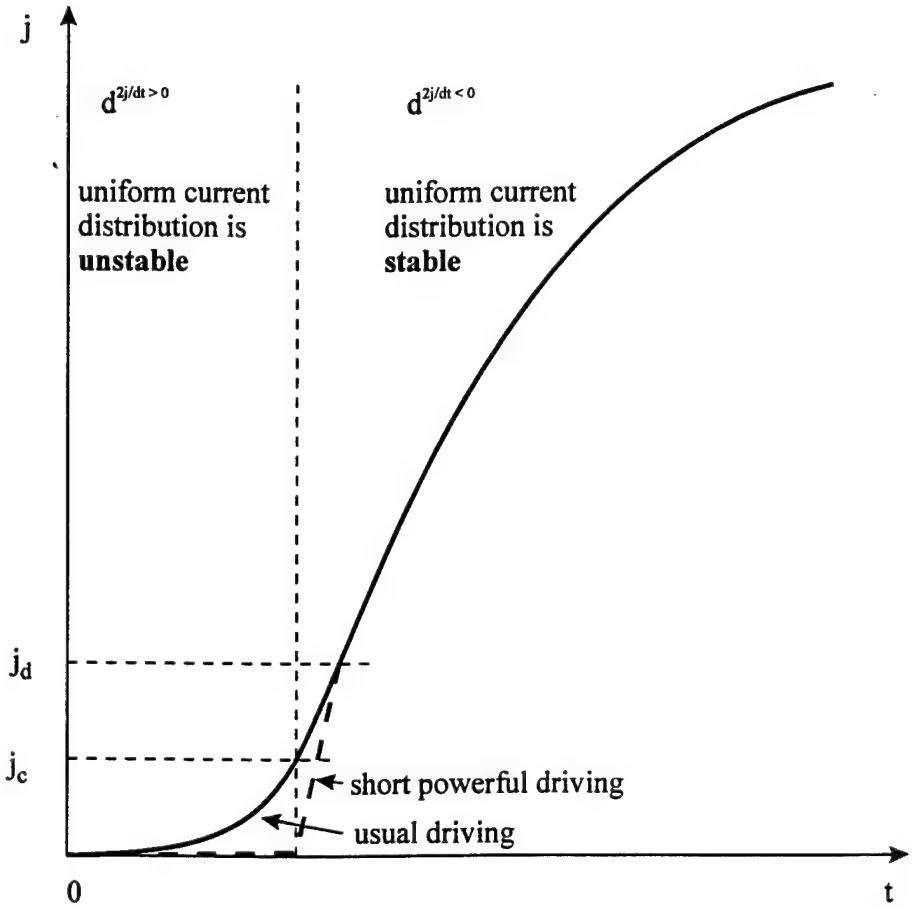
In spite of such perfect potentiality of thyristors the problem of fast switching of high currents remains. To get high dI/d factor it is necessary to increase the area of devices through which current flows. The main obstacle is dynamic filamentation of current in

turn-on process. This phenomena is typical for every device with positive (regenerative) feed-back.

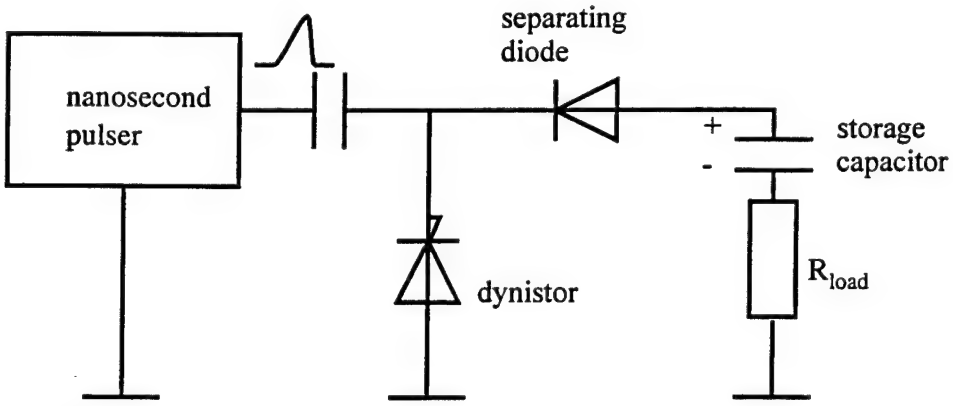
It was shown [6] that in thyristor like device such unstability of uniform current distribution appears in the turn-on processes when condition $d^2j/dt^2 > 0$ is fulfilled. This condition means: if in some small local area fluctuation with increased current density appears, the rate of current rise increases, so the current density increases again and so on. This local point current density "runs out" from a average current density in remained part of area (uniform current distribution is unstable). To suppress this unstability it is necessary to design $n^+p^+np^+$-structure in which the condition $d^2j/dt^2 < 0$ is fulfilled, that is the current rise is sublinear with time, at least at current levels exceeding some critical value j_c(Fig. 3.3). It is impossible to get from initial "off" state with $j \equiv 0$ to the stage where $d^2j/dt^2 < 0$ not crossing area where $\frac{d^2J}{dt^2} > 0$. Nevertheless if this crossing is placed in low current region, it is possible to "jump over" it by fast increase current density across all area up to needed value $j_d > j_k$ (Fig. 3.3).

3.2.1 Dynistors

Such "jump" is possible only in dynistors and may be forced by short $(2 \div 20ns$, high voltage pulse, applied to cathode and anode of a dynistor (Fig. 3.4). A separating diode prevents driving pulse from getting to the load. We have tested different types of tailor-made dynistors having *n - layer* width $W_n \approx 130\mu m$, 40 Ohmcm resistivity and *p+ - layer* (base) width $W_p \approx 10\mu m$. The dynistors held $(1 \div 1.2)\ kV$, have $15 \div 20n$ turn-on time. Applied driving pulse increases dynistor voltage over initial voltage U_0 of charged capacitor. If the length of driving pulse is less than dynistor turn-on time, the driving pulse "overvoltage" the dynistor above breakdown of collector p^+n-junction and initiate dynistor current j_d exceeding critical value j_k (Fig. 3.3). High electric field at *p+n- junction* causes impact ionisation, that gives fast rise to current. As was shown (2.17) impact ionisation could provide very high *dj/dt,* many orders of magnitude more than injection (2.8). To switch all dynistor area uniformly, driving current has to be approximately 10% of the maximum load current. With blocking voltage $U_0 = 1$ kV, 500 A were switched into the load for less than $20 \div 30n$. The blocking voltage may be easily increased by connecting several dynistors in series ("stacking" them). Two stacked dynistors switched 760 A at $U_0 = 2$ kV for the same time. Although dynistors appear to have considerable promise, they were not of considerable current use. The reason was - complicated driving systems. Now the use of step recovery devices makes



current density increase in thyristors ans dynistors



test bench for dynistors

Fig.3.4

generation of short driving pulses effective and simple. Nevertheless much research and design must be made before considerable use of dynistors.

Use of very short fronts (< 2 ns) allows to switch on dynistor in the delayed ionisation mode with turn-on times of **hundred picosecond.** The mode will be considered later in section devoted to picosecond devices.

3.2.2 Thyristors

Current limitation

Currently thyristors are of the most considerable use in power circuits. They need very simple driving circuits but their main disadvantage is relatively (in respect to transistors) low repetition rate.

Very cheap and simple 1 kV rated thyristor KU221 showed (when switched in *LC*-circuit $T_h = 160ns$, $\rho = 23\ Ohm$, $I_1/I_2 \approx 1,4$, $I_1 = 18A$) nearly the same figure as bipolar and FET transistors. But this thyristor, having the same as IRF840 case, showed the same I_1/I_2 = 1,4 factor up to more than I_1 = 100 A. Tested above transistors showed I_1/I_2>2 at currents exceeding 50 A. Our tailor-made for high *dI/dt* factor thyristors were tested in very low impedance *LC* circuit with U_0= 1,1 kV T_h= 300 ns, $C = 0,4$ nf and inductance determined only by wiring . The test showed I_1 = 3,3 kA, I_1/I_2= 2,8 and $dI/dt \approx$ 19 A/ns. The *dI/dt* advantages of thyristors are clearly seen, when they used in thyristor-diode closing switch (TDCS), briefly considered above (Fig. 1.7).

The operating mode of a thyristor in TDCS systems differs substantially from the operating mode in a linear modulator system, where the thyristor is switched on in series with the load resistance. In a normal modulator system, as the current increases through the thyristor, the voltage on it decreases, since the voltage drop on the load increases. At the end of the transition process, the current through the thyristor $I_{mH} = U_0/R_H$.

In a thyristor-diode switch, the voltage drop on the DSRD at the high conductivity stage is small, and when there is a large enough additional capacitor C_p, the increase in the current is limited by the charge Q_t, which was accumulated in the DSRD during the pumping time: the thyristor current increases when there is a constant voltage on it up to value I_m, which is determined by the condition $Q_t = \int_0^{Im} I(\tilde{t})d\tilde{t}$. Then, due to the superfast restoration of the voltage on the DSRD, the diode current breaks and switches to the load, while the voltage on the thyristor drops sharply. Such a system for increasing the current when there is a constant voltage and the external limitation of the charge has a number of peculiarities. The

concept of the "front for switching on the thyristor" loses its meaning, since the moment for the limitation of the current is controlled by an external circuit. The rate of increase of the current $I' = dI/d$ for the given voltage becomes the most important parameter.

Let us recall that the current which is cut off by the DSRD can easily be made as large as wanted: it is necessary only to increase its area (of course all limitation from chapter 2 are valid still). At the very same time, it is well known that increasing the area of the p^+--n--p--n^+ structure of the thyristors does not lead to a proportional increase in the rate of increase of the current due to the effect of the localisation of the current. A substantial increase in the area, and, consequently, in I' as well, turns out to be possible when switching on the thyristor using a overvoltage pulse along the anode . As was shown in [6], the presence of feedback between the current and the voltage in thyristors (due to the load reaction) contributes to the localisation of the current in the thyristor during the transition switching-on process. In a circuit where stored in DSRD charge is used to limit the thyristor current, such a coupling is absent, and the homogeneity of the distribution of the current during switching on can be improved.

The transition processes for switching on thyristors in the charge limitation mode have not been studied elsewhere. The maximum allowable values for I' given in the reference literature concern the system when the current is limited by an external circuit. These I' values can not even indicate the order of magnitude of the values in the case which interests us. Below are given some of the results of the study of modulator thyristors in a system for limiting the current by use of a stored charge. KU108 and KU221 thyristors were studied. As the experiments showed, these thyristors work very well in a pair with a DSRD.

Charge limitation

Thyristors were switched on according to the circuit shown in Fig. 1.7. Special attention was paid to decreasing the parasitic inductance of the additional capacitor C_p-- thyristor -- DSRD discharge loop, up to a value less than 50 nH. The current through the thyristor was recorded using a shunt; the temporal resolution was better than 10 ns. During the time for increasing the current until its breaking, the voltage on the thyristor, which is determined by the discharge of the additional capacitor and by the parasitic parameters of the circuit (inductance, active resistance), changed by no more than 10%. The current I_m, which was broken by the diode, is determined by the charge of the pumping pulse having duration $\tau_i = 0.5\mu$. The integral of the back current through DSRD during its flow time is equal to the charge Q_i which was introduced to the base of the diode during pumping.

In Fig.3.5 there are given the curves for the increase in the current, which were taken for KU108 at maximum currents 20 A and 140 A, and in table 3.1 there are given the averaged I' values which correspond to the voltage on the thyristor 1400 V. In the figure, for the sake of comparison, there are given the curves for the increase in the current when switching on a thyristor by the usual method (linear modulator) in a load resistance -storage capacitor- thyristor circuit for the very same maximum current I_m. It is evident that in the system for limiting the charge, the rate of increase in the current turns out to be higher than in the normal linear modulator system. The difference is the greatest when there are small currents -- five times less when $I_m = 20A$; when $I_m = 140A$, it is 2 times less.

The maximum rate of increase of the current increases almost linearly with an increase in the voltage on the structure which is evident from Fig.3.6.

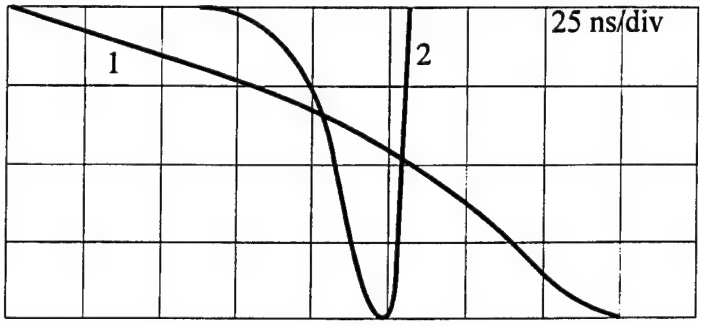
The increase in the rate of increase of the current in the charge limitation system is naturally connected with the unchangeability of the voltage on the structure, and, consequently with the strong field in the base layers and high velocity of transfer of carriers through them during the transition process.

Experiments on the study of the distribution of recombination radiation showed an increase in the switched - on area of the thyristor by 2-3 times in the charge limitation system. However, even when the control current is 15 A and when the front for its increase equals 20 ns, total switched area along the emitter - gating electrode boundary was less than 70%. Nevertheless, in operating modes with maximum speed values (I'_m), it is desirable to use control pulses with a short front (10-20 ns) and an amplitude equalling 10% and more of the maximum power current, which significantly surpasses the data book value. KU221 thyristors had $I'_m \approx 10^{10} A/s$, which is higher than that of KU108, even in the case of the smaller maximum working voltage which for KU108 thyristors reaches 1500 V, and for KU221 thyristors reaches 900 V.

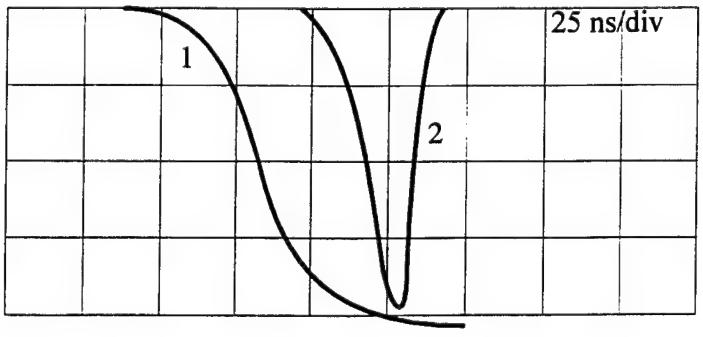
Table 3.1.

I_m, A	charge limitation	current limitation
21	$0.8*10^5$ A/s	$0.2*10^9$ A/s
70	$3.0*10^9$ A/s	$1.4*10^9$ A/s
140	$6.0*10^9$ A/s	$3.2*10^9$ A/s

a)



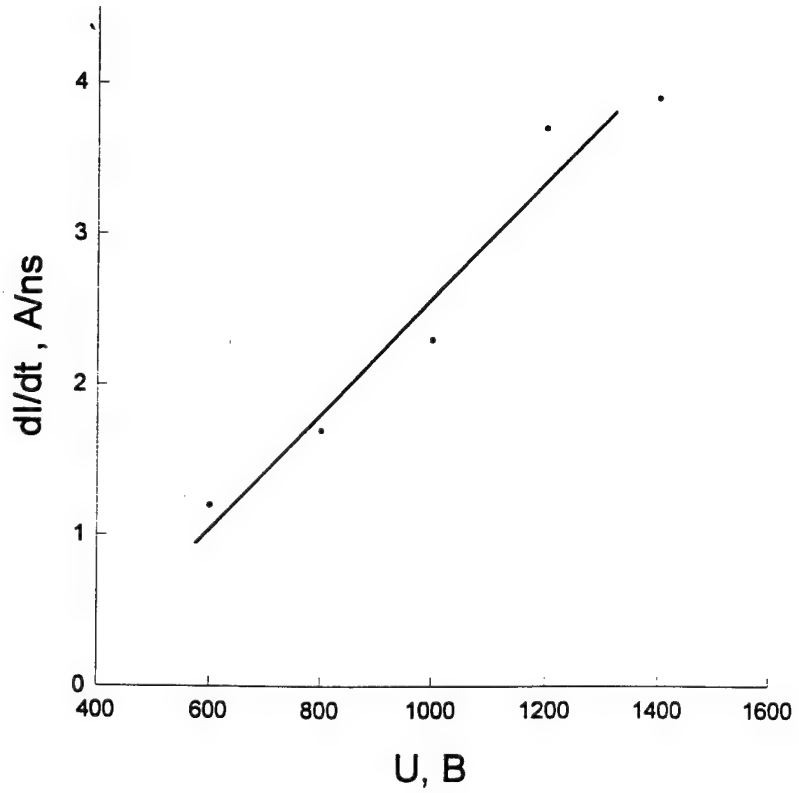
b)



1- current limitation
2- charge limitation

Current rise for KU108 with gating current 2A, maximum current 20A (a) and 140A (b).

Fig 2.5



dependance of dI/dt factor for KU108 on voltage

Fig.3.6

References

1. I.V.Grekhov, A.F.Kardo-Sysoev, "Subnanosecond current drops in delayed breakdown of silicon p-n junction". Sov.Tech.Phys.Lett., vol.5(1979), n.8, pp.395-396.

2. I.V.Grekhov, V.M.Efanov, A.F.Kardo-Sysoev, S.V.Shenderey, "Formation of high nanosecond voltage drop across semiconductor diode". Sov.Tech.Phys.Lett., vol.9(1983), n4.

3. I.V.Grekhov, "New principles of high power switching with semiconductor devices". Solid-State Electronics, vol.32(1989), n11, pp.923-930.

4. Yu.A.Kotov and all, "A novel nanosecond semiconductor opening switch for megavolt repetitive pulsed power technology". in Proc. 9th IEEE Pulse Power Conf.,1993

5. E.A.Panutin, I.G.Chashnikov."Current density distribution in fast thyristors". Sov.Radiotechnika and Electronica. v.23D(1978), n4, pp.883-886.

6. V.I.Brylevski, A.F.Kardo-Sysoev, I.G.Chashnikov, "Uniform current distribution unstability in fast semiconductor switch". Sov.Electronic.Tech., v.4(1985), n2, pp.48-52 (Russian).

7. A.F.Kardo-Sysoev and all, "Avalanch injection in high speed thyristors", IEEE trans. on electr.dev. ED-23(1976), n11, pp. 1208-1211.

Semiconductor Power Pulsers with less than 1 nanosecond front and up to 100 kV output

Report

Part 2

Pulse System Group Ltd.

1997

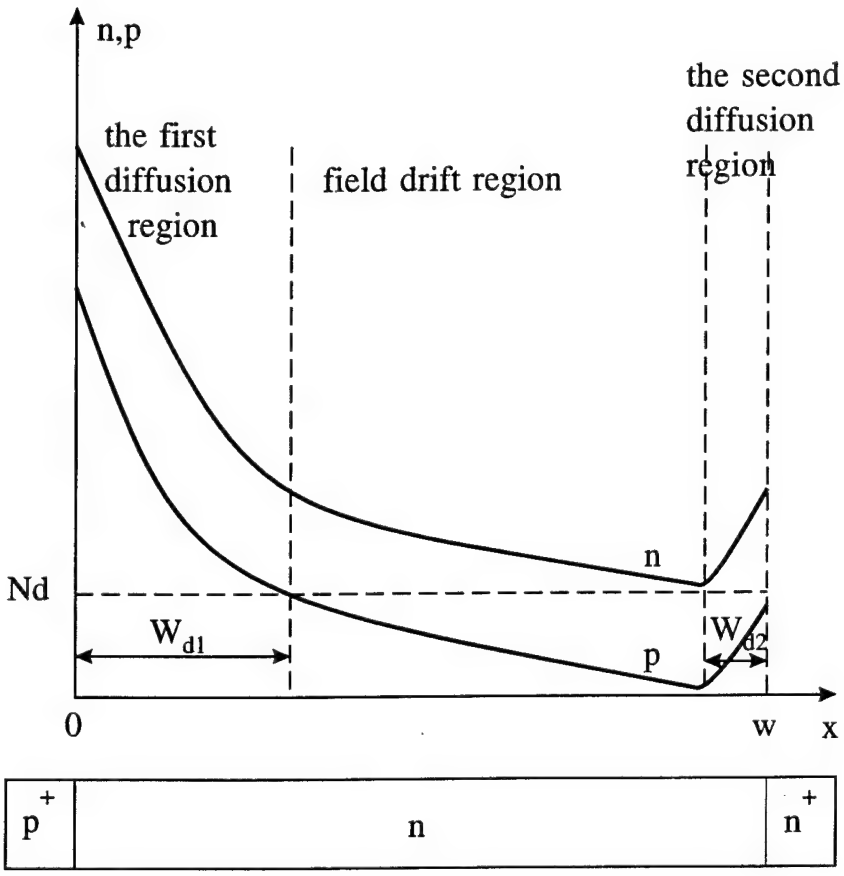


Fig.4.1
The distribution of the carrier concentration during injection in the diode.

4. Investigation of properties and limitations of nanosecond opening switches, based on step recovery of high voltage p-n junctions

4.1 General consideration

We have covered the basics of the drift step recovery diodes in section 1. To estimate the potentiality of the devices it is necessary to consider the physics of their operation. No work on the devices has been published in the West. Currently only in Russia there are several groups working on drift step recovery devices. The most of them are involved in DSRD pulsers design. Only one group which is with Mega Pulse and PTI is involved in investigation of the physics of DSRDs and the device R&D. The following consideration of the physics is based on the work of this Mega Pulse group. Attention now focuses on estimation of the main relations between the devices technological parameters (n-layer width, doping level and so on) and operational parameters such as optimal current, switching time, voltage drop and others.

To begin with we will make simplified estimation of the upper limit for the main parameter dU/dt, then to support the conclusions, more detail theory will be presented as a supplement.

In order to explain the effect of super-fast restoration in high-voltage diodes, the "classical" well known pattern of the processes must be expanded considerably. The theory of the effect is stated below, and we will now examine its basic properties at the qualitative level.

In the case of the sharp switching on of a forward current pulse I_+ with duration τ_+ in the p^+-n-n^+ diode nearby the p^+ --n junction during the current flow time, a diffusion layer of plasma is formed (figs.1.1, 4.1) with a characteristic size $W_{d1} \approx \sqrt{D_p \tau_p}$ and a maximum concentration by the p^+ --n junction of $n_m \approx j_+ \tau_+^{1/2}/(qD_p^{1/2})$.

The small duration of the current pulse τ_+ as compared with the life time of the minority carriers τ_p ($\tau_+ << \tau_p$) is a very important condition in the pattern being described. Beyond the border of the diffusion layer, where $x > W_{d1}$, the concentration of the minority carriers is small, and their transfer is determined by the drift in the electrical field. When there is a sufficiently large charge which has passed during time τ_+, a second narrow enriched diffusion layer with size W_{d2} can be formed at the second boundary by the n-n^+ junction. The size of the second diffusion area and the concentration accumulated in it turn out to be less than in the first diffusion area because of the fact that the holes move over to the second area with a delay for the time of their flight through the n-base.

It can be shown that a sharply nonhomogeneous distribution which is close to that mentioned above is also obtained in structures whose minority carriers have a small lifetime. In this case, in the event of all of the arguments above, it is necessary to substitute τ_+ for τ_p.

Thus, when there is a decrease in τ_+, the size of the enriched layer by the p^+--n junction can always be made smaller than the final size of the SCR when there is complete restoration of the voltage. In our case, $W_{d1} \approx 15$ μ when $\tau_+ = 2 \cdot 10^{-7}$ s and $W_{SCR} \approx 70$ μ when $U \approx 1$ kV. It is precisely this characteristic which provides the super-fast restoration. Due to the small size of W_{d1}, the voltage drop on the SCR after its formation is small during the entire time while its boundary remains inside W_{d1}. As the SCR boundary passes through the enriched plasma layer, its thickness decreases. After the boundary leaves for the area with a low concentration, the rate of expansion of the SCR sharply increases, and as a consequence of this the increase in the voltage on the SCR also accelerates. As further analysis shows, by the

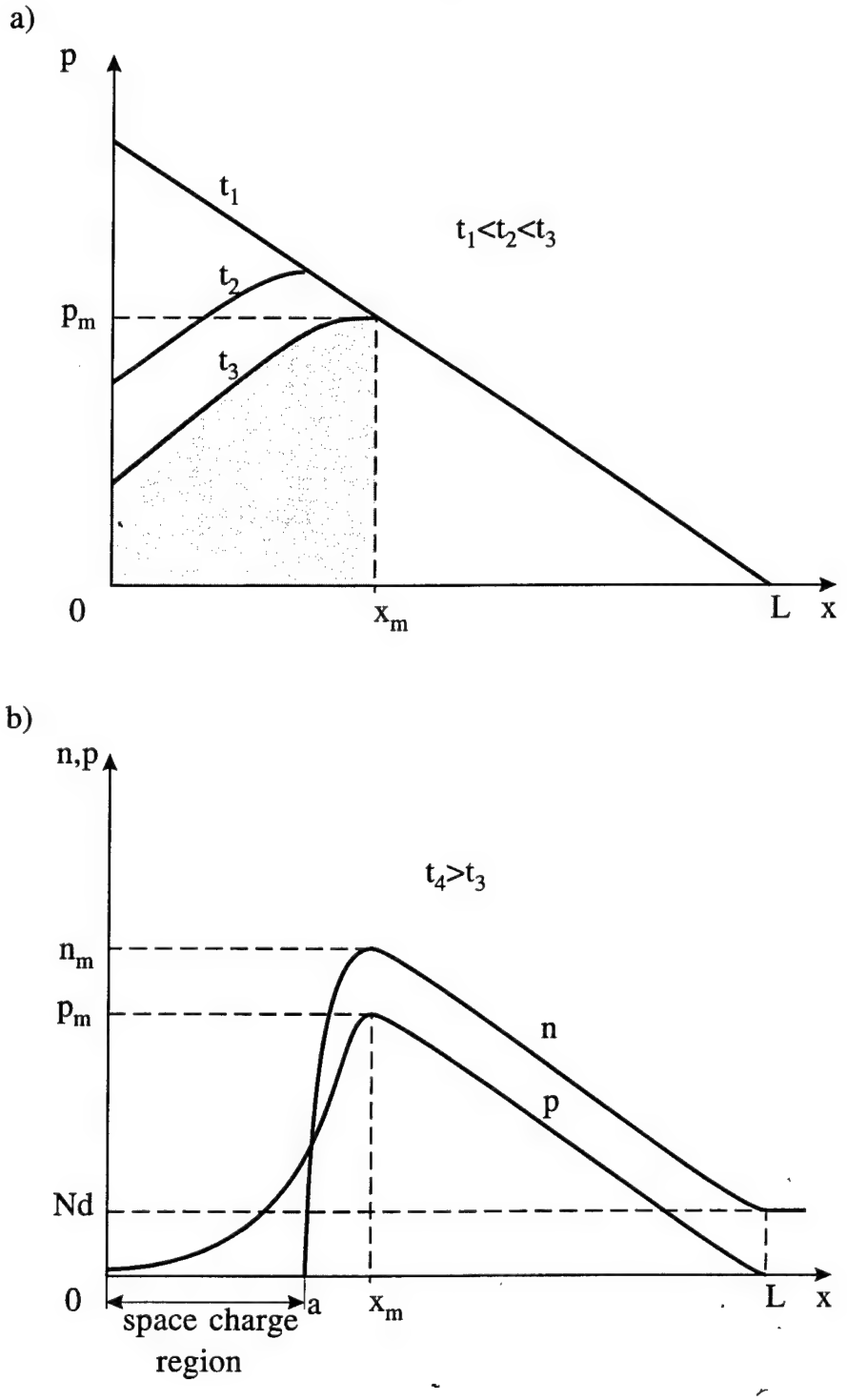


Fig.4.2
The restoration of the space charge region
by the p^+n junction

time the enriched layer disappears, all of the minority carriers also turn out to be pulled out of the drift region ($x > W_{dl}$), and the conductivity current in the SCR is broken.

From what has been stated above, it is possible to immediately obtain elementary evaluations for the rate of growth of the voltage on the diode in the process of restoration during the passage of reverse current with density j_-. The intensity of the field in a quasineutral region is inversely proportional to the carrier concentration:

$$E \approx \frac{j_-}{2q\mu}, \tag{4.1}$$

where μ is the mobility, μ of the electrons for simplicity's sake is assumed to be equal to μ of holes.

As was noted (see fig. 4.2), by the p^+--n junction there arises an area with a strong plasma concentration backward gradient, which will contribute to the move of the holes to the p^+-layer, and will prevent the leakage of electrons from the p^+--n junction. With time, the plasma concentration in the area drops, while, in agreement with expression (4.1), the field intensity begins to increase. It is obvious that at the point of the concentration maximum at which $dn/dx = 0$, the electrons move away from the p^+ layer to the n layer with a speed which depends on the intensity of the electrical field created by the passing current:

$$v_m = \mu E = \frac{j_-}{2qn_m}. \tag{4.2}$$

After the time during which an electrical charge equaling the entire charge of the electrons in the dotted area of fig. 4.2 passes through the external circuit, the concentration of electrons by the p^+--n junction reaches zero. The neutrality near the junction is disrupted, and an area of positive space charge, determined by the ionized impurities with a concentration of N_d and the holes, arises. The field intensity in the SCR and the voltage drop on it U_{SCR} begin to increase. The boundary (a thickness on the order of the Debye radius) between the SCR and the neutral plasma moves with the speed of the electrons' departure from the boundary. The rate of movement of the boundary does not coincide with the speed of the point of the concentration maximum (this problem will be examined later in more detail). However, in the first approximation, due to the retention of the concentration gradient, they can be considered to be close. Then, taking expression (4.2) into consideration, the rate of growth of the voltage on the SCR is

$$U'_{SCR} = \frac{dU_{SCR}}{dt} = \frac{(qN_d + j_-/v_s)av_m}{\varepsilon} = \frac{(qN_d + j_-/v_s)aj_-}{2\varepsilon q n_m}, \tag{4.3}$$

where a is the coordinate of the SCR boundary; n_m is the concentration of the plasma at this boundary.

Upon the conclusion of this expression for U'_{SCR}, it is supposed that the field intensity in a large part of the SCR exceeds the value $E_s = 10^4$ V/cm, which corresponds to the saturation of the drift velocity of the charge carriers. Expression (4.3) shows that the maximal for $U'_{SCR}(U'_m)$ value is determined only by the constants of the material and equals

$$U'_m = E_m v_s, \tag{4.4}$$

where E_m is the maximum value of the field in an SCR limited by an avalanche breakdown.

For silicon, $E_m \approx 2 \cdot 10^5$ V/cm, $v_s \approx 10^7$ cm/s, and $U'_m \approx 2 \cdot 10^{12}$ V/s. This means that it is possible in principle, using diodes, to shape a change of voltage with an amplitude of close to 2 kV during time τ_ϕ of approximately nanoseconds, which is extremely attractive for nanosecond pulse technology.

The ideas which have been formulated by the present time on the dynamics of double injection processes and the dispersal of plasma in semiconductor devices with p--n junctions will be expounded in detail in the next supplementary

4.2 The theory of nonstationary injection processes and the extraction of carriers in a semiconductor

4.2.1 Double Injection

The processes which take place in a semiconductor are described at the hydrodynamic level by a system of equations which consists of the continuity equations (2.1) and (2.2) and the Poisson equation (2.3), supplemented by the current equations

$$j_p = \mu_p p E - D_p p_x;$$
$$j_n = \mu_n n E + D_n n_x; \qquad (4.5)$$

$$j = j_p + j_n + \varepsilon E_t. \qquad (4.6)$$

The t and χ indices designate the derivatives for time and the coordinate. Equation (4.6) is a consequence of the preceding equations, nevertheless it is convenient to write it explicitly.

In agreement with the our task, we will only be interested in the fast processes (less than 10^{-6} s); therefore we will eliminate recombination and thermal processes from the discussion.

As is well known, the kinetic coefficients (mobility, the diffusion coefficient) depend on the field intensity and the carrier concentration. In silicon, the effect of the electron-hole scattering which decreases mobility begins to have a noticeable influence when the concentration is higher than 10^{17} cm^{-3}. When $E_s \approx 10^4$ V/cm, the drift speed reaches the threshold saturation value 10^7 cm/s for electrons and $v_s \approx 0.6 \cdot 10^7$ cm/s for holes, which is equivalent to a decrease in mobility with an amplification of the field; $\mu \approx E^{-1}$. Let us first examine only relatively weak fields ($E < E_s$), for which the mobility can be considered to be constant. In fast processes, the concentration of carriers can not reach the large values which are characteristic for a stationary state when there is a high current density and when the large concentration effect begins to have an influence. Therefore, we will disregard the influence of the concentration on the kinetic coefficients. For the sake of being definite, we will examine an n-type semiconductor with thickness W, which is homogeneously doped with a donor impurities having concentration N_d. Through the semiconductor a current flows with a characteristic density j_c. The initial system can, using the exclusion of j_p and j_n from expressions (4.1) - (4.3) and (4.5) and (4.6), be given in the following dimensionless form:

$$j = \tilde{E}\tilde{Q} - \xi\tfrac{1-b}{1+b}\tilde{Q}_x + \eta\tilde{E}_t - \tfrac{2b}{1+b}\xi\eta\tilde{E}_{xx}; \qquad (4.7)$$

$$\tilde{Q}_t = b\tilde{E}_x - b\eta(\tilde{E}\tilde{E}_x)_x - (b-1)\eta\tilde{E}_{xt} + a\xi\tilde{Q}_{xx} + \tfrac{3b(1-b)}{1+b}\xi\eta\tilde{E}_{xxx}, \qquad (4.8)$$

where $\tilde{Q} = \tilde{p} + b\tilde{n}$;

$$a = \tfrac{2(1-b+b^2)}{1+b}; \quad \xi = \tfrac{kT}{qE_u w}; \quad \eta = \tfrac{\varepsilon E_u}{qN_d w}.$$

As the units the following values are given here: current density $j_u = j_m$; field intensity $E_u = j_u/(q\mu_p N_d)$; concentration $n_u = N_d$; potential $U_u = E_u w$; size $\chi_u = w$ (the thickness); time $t_u = w/\mu_p E_u$ (flight time for the holes).

The system of equations (4.7) and (4.8) contains two dimensionless parameters ξ and η. The ξ parameter is the ratio of the thermal potential $U_T = kT/q$ to the potential applied to the semiconductor: $\xi = U_T/U_u$. The η parameter is the ratio of the Maxwell relaxation time $\tau_\mu = \varepsilon/(q\mu E_u)$ to the flight time of the majority carriers in characteristic field $\tau_E = w/(\mu_n E_u)$; $\eta = \tau_\mu/\tau_E$. It is easy to show that the τ_μ/τ_E ratio is also identical to the ratio of the thickness of the SCR in a semiconductor $w_{scr} = \sqrt{2\varepsilon U_u/(qN_d)}$, which arises when voltage U_u is supplied to it (the space charge is determined only by the ionized impurity), to the thickness of the semiconductor.

For a complete determination of the (4.7) and (4.8) equation system, it is necessary to give four additional conditions, since it is a fourth order system and one additional equation, since the system contains three unknowns. From the physical posing of the problem it is clear that this equation should connect the current $I = jS$ (S is the area) with the complete drop of the voltage on the semiconductor $U = \int Edx$ through the external circuit parameters. From the Kirchgoff law for a purely active circuit, it follows that

$$U = U_0 - R_l I, \qquad (4.9)$$

where U_0 is the voltage of the external generator; R_l is the resistance of the load.

In addition to the time condition, equation (4.9) also gives the condition for the boundaries; therefore it is necessary to give four more boundary conditions.

At the very same time that the initial condition is obvious enough for each specific problem, the problem of the boundary conditions requires special examination. We will first evaluate the characteristic parameters of the problem. For the conditions given above, during which the super-fast restoration effect arises, and has been investigated experimentally the main parameters have the following values: $W \approx 10^{-2}$ cm; $N_d \approx 10^{14}$ cm^{-3}; $j_u \approx 10$ A/cm^2; $\tau_M \approx 10^{-10}$ s; $\tau_E = 10^{-8}$ s; $U_u \approx 10$ V; $U_l = 2.5 \cdot 10^{-2}$ V; $\xi \approx 2.5 \cdot 10^{-3}$; $\eta \approx 10^{-2}$.

We will note that the potential U_u corresponds to the conducting phases of the operation of the diode.

So, we obtain the fact that in the given case ξ and η are small parameters (multipliers) before older derivatives in equations (4.7) and (4.8). Therefore, disregarding the term which contains the small parameter changes the order of the system and, consequently, the number of additional conditions as well; that is, equations (4.7) and (4.8) belong to the class of so-called singularly perturbed equations. The presence of the small parameter for the older derivative signifies the appearance of the so-called boundary layer; that is, a narrow area where the variable changes very quickly. For singularly perturbed systems, computing procedures is known which makes it possible to construct asymptotic expansions for the small parameter.

The degenerated system obtained from equations (4.7) and (4.8) under conditions $\xi \to 0$ and $\eta \to 0$:

$$\tilde{Q}_l = b\tilde{E}_x; \qquad j = \tilde{E}\tilde{Q} \qquad (4.10)$$

is first order; therefore, in the initial system no less than three boundary layers should be expected, and the construction of an asymptotic solution is practically impossible. When simplifying the system, we use a number of physical considerations, for which we first specify the problem in agreement with the conditions for the operation of the diodes.

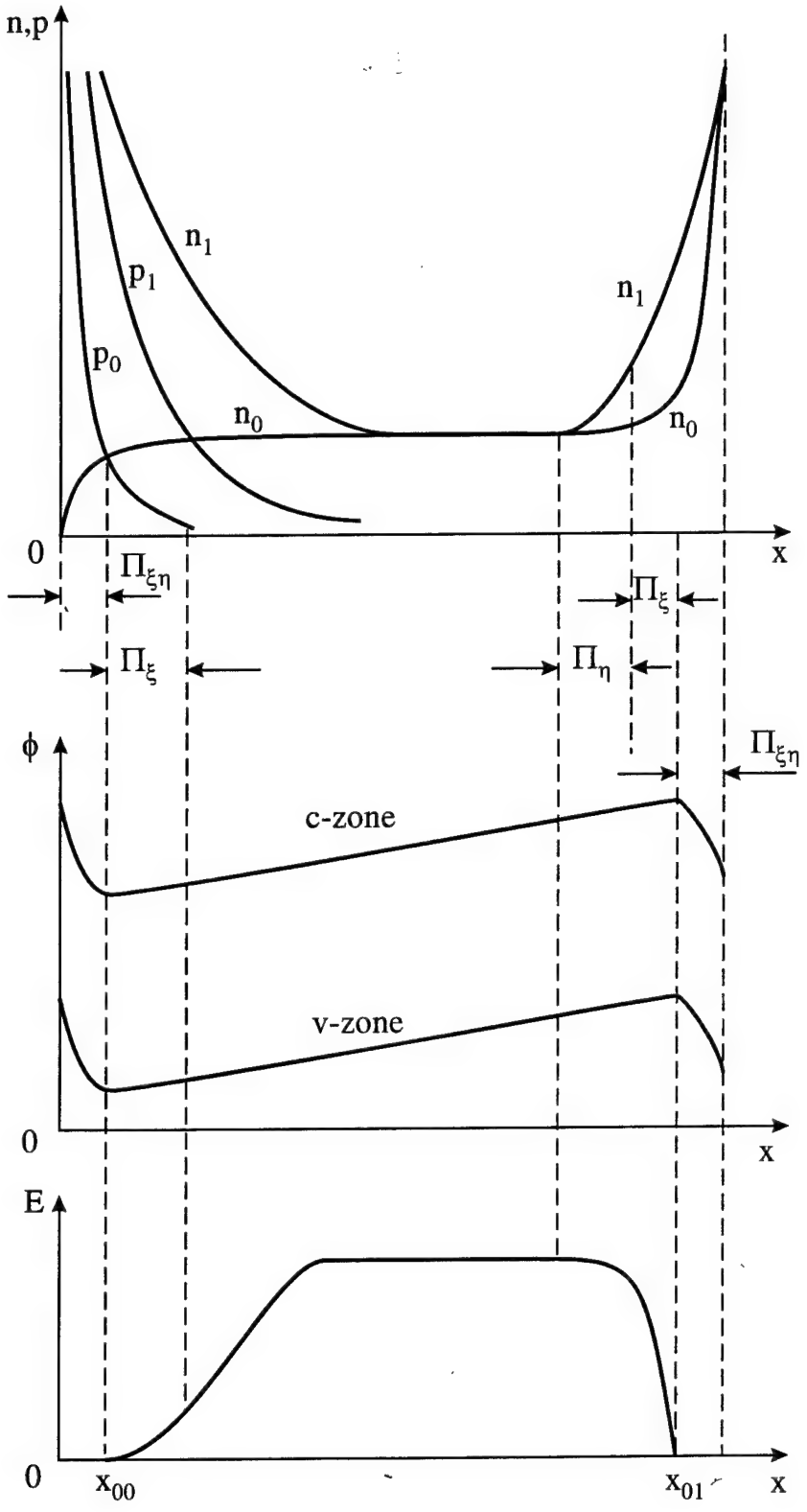


Fig.4.3
The disposition of the boundary layers in a p^+nn^+ structure in the case of the injection of carriers.
ϕ is the potential.

Let us examine a p^+ -- n -- n^+ type of structure; that is, the n layer of the semiconductor is placed between the p^+ -- n and n^+ -- n junctions. We will accept that during a time which is small as compared with the characteristic flight time $\tau_E \approx 10^{-8}$ s, but large in comparison with the Maxwell relaxation time $\tau_M \approx 10^{-10}$ s, a constant current with a density of j_+ is established. When there is such formulation of the problem, equation (4.9) turns out to be superfluous. The value of the $\xi\eta$ parameter which stands by the older derivative E_{xxx} in equations (4.7) and (4.8), as is easily shown, is equal to the square of the ratio of the Debye radius to the thickness: $\xi\eta = U_T\varepsilon/(qN_dW^2)$ and equals 10^{-4} with respect to order of magnitude. Thus, the characteristic size of the boundary layer corresponding to the $\xi\eta$ parameter is equal to the Debye radius 10^{-4} cm, and its time lag is equal to the Maxwell relaxation time (approximately 10^{-10} s). Therefore, it is possible to consider the processes in this layer to be stationary with respect to the processes for the transfer through the n-area with a time of approximately 10^{-8} s.

As was shown in reference [8], the boundary layer which corresponds to the $\xi\eta$ parameter is the area of the potential barrier by the p -- n junction. In our case, there will be two such layers: by the p^+ -- n and n -- n+ junctions (fig. 4.3). In the case of supplying a current in the forward direction to a p^+ -- n -- n^+ structure, the electrons in the n region move under the effect of a field to the p^+ layer, where they stop in front of the barrier by the p^+ -- n junction and are accumulated. The holes entering the n region from the p^+ layer compensate for the space charge of the electrons being accumulated in front of the barrier. Since the field intensity is identically equal to zero by the border of the barrier at point χ_{00} of the twist extremal point in the potential curve, then due to the smallness of the space charge (due to the compensation for the charge of the holes and electrons in the area by point χ_{00}) when $\chi > \chi_{00}$, a weak field region will exist in which the electrons and holes are carried by diffusion -- the Π_ξ layer. All because of the presence of the space charge, although it is small, at a great enough distance from point χ_{00}, the field is intensified so that the field transfer becomes a determining factor.

For time intervals which are smaller than the hole's flight time through the n layer ($t < t_w$), by the n^+ -- n junction the entire current is transferred only by electrons entering the n layer from the n^+ layer. At the point that the curve for the potential χ_{01} twist, just like at the χ_{00} point, the transfer is purely a diffusion transfer; however, due to the presence of the uncompensated space charge of the electrons at a distance from this point ($\chi < \chi_{01}$), the field intensity increases very quickly, and the field transfer becomes a determining factor. In other words, a layer Π_η of the space charge of the electrons with a drift transfer mechanism appears, as a consequence of which the size of the diffusion layer Π_ξ by the n^+-- n junction turns out to be many times less than the size of the Π_ξ layer by the p^+-- p junction.

After the flight time t_w, the holes approach the barrier by the n^+-- n junction, and, accumulating in front of it, they begin to compensate for the space charge of the electrons. The characteristic time expended for this compensation can be easily evaluated as follows: it is close to the Maxwell time τ_M. After the compensation, the Π_η layer disappears. As a result, the field gradient in the area of point χ_{01} begins to decrease, and the diffusion layer begins to expand, and this occurs in a similar fashion to the way this occurred by the p^+--n junction. After a sufficiently large time interval, the patterns for the distribution of the fields and concentrations by the p^+-- n and n^+--n junctions become almost identical (with an accuracy up to the ratios for the carrier mobility and the leakage of carriers across the barriers).

In summing up what has been stated, it is possible to qualitatively represent the following pattern for the disposition of the $\prod$ layers (Fig.4.3):

two barrier $\Pi_{\xi\eta}$ layers (by the p^+-- n and n^+--n junctions) corresponding to parameter $\xi\eta$, with a characteristic relative size $(\xi\eta)^{1/2}$ and a stationary state establishment time η;

two diffusion layers Π_ξ corresponding to parameter ξ and bordering with the barrier layers $\Pi_{\xi\eta}$; the sizes of the Π_η layer by the p^+-- n and n^+--n junctions can differ greatly from one another;

one Π_η layer bordering the diffusion layer by the n^+-- n junction;

a regular layer with a field transfer placed between the layer Π_η (by the n^+--n junction) and the Π_ξ layer (by the p^+--p junction) and described by equation (4.10), which does not contain small parameters.

The thickness of the barrier layer $\Pi_{\xi\eta}$ is less than 10^{-2}, and the time lag is $\eta \approx 10^{-2}$. Therefore, the barrier layer can be represented as an infinitely thin layer, or wall, at which both the field and the concentration are sharply changed they undergo a break. For the very same reasons (small thickness and lag), it is clear that in the $\Pi_{\xi\eta}$ layer the continuity of the carrier flows should be retained; that is, the relative fraction of the flow going to change the full number of carriers in the layer is small. Hereafter, we will completely exclude the $\Pi_{\xi\eta}$ layers from our consideration, after replacing them with a break at which the flows of the carriers will be given. The simplified ($\xi\eta = 0$) system of equations (4.7) and (4.8) corresponding to this will have only second order.

It has already turned out to be possible to apply the procedure for searching for asymptotic expansion [9] to this simplified system, which contains two small parameters ξ and η; that is, searching for the least degenerated forms of shortening the initial system, which correspond to the boundary layers. Let us discuss layer Π_η. The solution for it should be meshed together with the solution for the regular layer; therefore, in it the characteristic values for the field intensity and the concentration are close to one. We will increase the scale for the length $\chi = (\tilde{\chi}_\eta - \tilde{\chi})/\eta^\lambda$, where $\tilde{\chi}_\eta$ is the boundary between layers Π_η and Π_ξ, while $\lambda > 0$ is a certain arbitrary number which should be defined. The shortened equation (4.8) will have the following appearance:

$$\eta\tilde{Q}_t = a\frac{\xi}{\eta}\tilde{Q}_{\chi\chi} - b\tilde{E}_\chi - b^{1-\lambda}\eta(\tilde{E}\bar{E}_\chi)_\chi + (b-1)\eta\tilde{E}_\chi. \qquad (4.11)$$

If $\lambda = 1$ is accepted , it is easy to check by the substitution that when $\eta \to 0$ in equation (2.11), the largest number of components is retained. In the area of parameters which interests us, as was shown earlier, $\xi/\eta < 1$; therefore, when $\eta \to 0$ and $\lambda=1$, we get an equation (characteristic expansion) which described the distribution of the field in the Π_η-layer:

$$\bar{E}_\chi + \frac{d}{d\chi}(\bar{E}\bar{E}_\chi) = 0. \qquad (4.12)$$

where the bar signifies belonging to the boundary layer Π_η.

Evidently, in the expansion there is no derivative for time. Physically this means that the lag of the Π_η layer is very small (approximatēly equal to τ_M) and the distribution of the field in it is quasistationary with respect to the slow processes in the regular layer. Additional conditions for equation (4.12)(4.12) should correspond to the conditions for meshing the solutions for the regular and diffusion layers

$$\chi = \chi_\eta \quad or \quad \chi = 0; \qquad \bar{E} = \bar{E}_\eta;$$

$$\chi \to \infty; \qquad \overline{E} = \overline{E}_l,$$

where $\overline{E}_1$ is the intensity of the field at the boundary of the regular layer with the Π_η layer; E_η is the intensity of the field at the boundary of the diffusion layer with the Π_η layer.

When integrating equation (4.12) under conditions at the boundary, we get a transcendent equation which describes the distribution of the field in the layer:

$$\chi = \overline{E}_\eta - \overline{E} - \overline{E}_l \ln \frac{\overline{E}_l - \overline{E}}{\overline{E}_l - \overline{E}_\eta}. \qquad (4.13)$$

It follows from here that when $\overline{E}_1 >> \overline{E}_\eta$, the distribution of the field is insensitive to the $\overline{E}_\eta$ value, and $\overline{E}_\eta = 0$ can be assumed. The characteristic size of the Π_η layer is η or in the dimension form $\varepsilon j_u/(q\mu_p N_d^2)$; that is, layer Π_η has a size equal to the distance which an electron flies through during Maxwell time . In the cases which interest us, the size of the Π_η layer equals approximately 10^{-4} cm (much less than W). Due to the small lag (τ_M) and the thickness, the Π_η layer, just like the $\Pi_{\xi\eta}$ layers, can be replaced by a wall with the break on it of the field and the concentration, but with the retention of the flows. Let us recall that after time τ_M, after the approach of the holes to the Π_η layer, this layer disappears.

When finding the expansion for diffusion layers, we cannot be limited by the extension of the coordinate with respect to the power of the ξ parameter, as this was done in the preceding case. The fact is that the characteristic concentration and the intensity of the field in these layers are incomparable with the characteristic values in the regular layer, where they are close to one. Therefore, we are obligated to change the scale of the dependent variables as well (to "compress" the field intensity and to "extend" the concentration):

$$\tilde{\chi} = \xi^\lambda \chi; \qquad \tilde{Q} = \overline{Q}\xi^m; \qquad \tilde{E} = \overline{E}\xi^k, \qquad (4.14)$$

where $\chi, \overline{Q}, \overline{E}$ are new meanings for the corresponding values; λ, $m < 0$, $k > 0$ are constants which must be found.

When substituting expression (4.14) into equations (4.7) and (4.8), we find that as was done similarly to this previously, the expansions for the diffusion layers when the values of the constants $m = -1/5$, $k = 1/5$, $\lambda = 2/5$:

$$\overline{Q}_t = b\overline{E}_\chi + a\overline{Q}_{\chi\chi}; \qquad j = \overline{Q}\overline{E}. \qquad (4.15)$$

It should be noted that from the expansion for Π_ξ the term $[(1-b)/(1+b)]Q_x$, which takes the Dember effect into consideration, is automatically excluded. Therefore, when examining the diffusion layer, there is no need to specially rule out the given layer (assuming $b = 1$), as was done, for example, in reference [10]. Furthermore, it is also clear from the physical considerations that the influence of the Dember effect is small.

The solution to system (4.15) should be meshed for the left diffusion layer Π_ξ with the Π_η layer. In addition, conditions should be given for each boundary of these layers with the p^+ -- n and n^+ -- n junctions: in our case, the carrier currents. As follows from a comparison of systems (4.10) and (4.15), the equation for the diffusion boundary layer also includes the equation for the regular layer. Therefore, the solution of system (4.15) for the boundary layer by means of the simple transformation of the scales could have been extended for the regular layer. However, an analytical solution to equations (4.15) is impossible. At the very same time, it is clear that in the area of the points for the zero field intensity χ_{00} and χ_{0P} the term in

system (4.15) which takes diffusion into consideration is a decisive one, and the distribution of the concentration can be described in the following manner:

$$\tilde{Q}_t = a\xi\tilde{Q}_{\chi\chi} \qquad (4.16)$$

Here we have returned to the initial scale [see equation (4.8)]. However, it should be recalled that this shortened equation is already impossible to correctly mesh (within the framework of the theory of asymptotic expansion, as was done for the Π_η layer) with the regular part (4.10). The meshing is possible only again through the intermediate region (4.15).

The incorrectness of the use of only equation (4.16),as was widely accepted (see, for example, reference [10]) for the description of processes which occur when there is a high injection level in the interval $\chi_M > \sqrt{Dt_M}$, where t_M is the time scale, is already clear now.

In experiments, integral values --- voltage and complete current -- are determined. Based on what has been stated, it is possible to evaluate the contribution to the complete voltage drop on a structure which is dictated by each boundary layer (relative numbers): the regular region $U \approx 1$; the space charge layer $\Pi\eta \to U_\eta \approx \eta << 1$; the diffusion layer $\Pi_\xi \to U_\xi \approx \xi^{k+\lambda} = \xi^{3/5} << 1$ Thus, the total contribution of all of the boundary layers does not reach ten percent of the contribution of the regular layer. Therefore, in order to compare the theory and experiment, it is sufficient to obtain a qualitative pattern for boundary layers. Obviously, due to the continuos accumulation of carriers in the diffusion layer, the diffusion layer should be expanded. In the layer by the p^+-- n junction, this process begins immediately after the current is switched on, and in the layer by the n^+-- n junction the beginning of the process of the accumulation and expansion lags for the flight time of the holes through the regular layer. From the appearance of equation (4.16), it follows that the characteristic size of the diffusion layers (L) should increase in time like $L \approx \sqrt{D_p t}$, where t is the time calculated from the moment expansion begins.

The boundary condition for equation (4.16) is obvious: at the barrier of the p^+- n junction, the electron current equals zero (when the injection coefficient $\gamma_p = j_p/j = 1$), from which it is easy to derive the well known equation $j = -2qD_p dp/d\chi$ or in the dimensionless form $j = -\xi(1+b)\bar{Q}_\chi/b^2$. From this boundary equation it follows that, in a certain region near the boundary, the change in the concentration is linear in character. Using equation (4.16), it is possible to show that the linear character will be retained in the area where the condition $\tilde{\chi}^2/(2\xi t) << 1$ is fulfilled, and in that part of the diffusion layer where $\tilde{\chi}^2/(2\xi t) >> 1$, the solution to equation (4.16) has the following appearance:

$$\tilde{Q} \approx \tilde{\chi}^k \exp(-\tfrac{\tilde{\chi}^2}{4\xi at}), \qquad (4.17)$$

where k and a are constants which should be defined.

It is possible to obtain a completely solution for all $\tilde{\chi}^2/(2\xi t)$ values while meshing the linear part with the exponential part [expression (4.17)] at a certain point $\tilde{\chi}^2/(2\xi t) = 1$. The k and a constants can be chosen from the condition for the continuity of the concentration and its derivative at the meshing site, and from the condition for the minimizing of the imbalance which arises when equation (4.17) is substituted into the initial equation (4.16).

However, such a significant complication to the solution, while impeding the qualitative analysis of the processes, by virtue of what has been stated above will have little effect on the numerical processes, and all the same will not allow the correct meshing of the

solutions for the diffusion and field regions. Let us stress yet again that the intermediate region of the meshing can be obtained only by the solution to system (4.15), which, in its turn, we have found by numerical methods, which will make it possible to evaluate the contribution of the intermediate region of the meshing to the error in the analytical description given here.

Therefore, we will then use the simplest "linear" approach.

At the boundary between the diffusion layer and the regular layer, the ratio of the hole current to the complete current, is determined by the ratio in the regular layer; that is

$$\frac{j_p}{j} = \frac{p}{(1+b)p+bN_d}.$$

The flow of holes entering the diffusion layer by the p^+- n junction through the barrier has a density $j_p = \gamma_p j$, where γ_p is the barrier injection coefficient; the density of the outgoing flow is determined by the ratio obtained above, which for the high injection ($p >> N_d$) cases which are of practical interest is simplified: $j_p = j/(1+b)$. These ratios make it possible to make an easy determination of the complete number of holes accumulated in the diffusion layer by the p^+- n junction:

$$P_p = (\gamma_p - \frac{1}{1+b})P_j, \tag{4.18}$$

where $P_j = \int_0^t jd$ is the complete number of carriers which passed through the external circuit.

Upon solving equation (4.16), it is possible to accept the following (in dimensional form)

$$p = p_m(1 - \frac{\chi}{L}), \tag{4.19}$$

(where $p_m = j_+L/(2qD)$; $L = \sqrt{2Dt}$; $D = 2D_p b/(1+b)$ is the effective diffusion coefficient), as the first linear approximation for the region $\tilde{\chi}^2/(2\xi t) < 1$, taking into consideration the boundary condition when $\tilde{\chi} = 0$, and taking into consideration expression (4.18).

Upon solving equation (4.10) for a regular layer, which corresponds to the known equation for bipolar drift [11], we will use the approach described in the reference. It has been noted that, at the boundary with the diffusion layer, the field is weaker than at the opposite end of the layer, and the solution is insensitive to the value of the field intensity at the boundary which therefore can be assumed to be equal to zero. Thus, in the formulation of the problem, there is no natural length scale up to the moment of the hole's flight through the n layer. In this case, the problem is automodeled. When introducing the new variable $y = \tilde{x}/\tilde{t}$ and substituting it into system (4.10), we get when $j_+ = \text{constant}$

$$\tilde{E} = \sqrt{\frac{\tilde{\chi}}{bt}}; \quad \tilde{Q} = \sqrt{\frac{bt}{\tilde{\chi}}}. \tag{4.20}$$

The solution to (4.20) automatically satisfies the condition $\tilde{\chi} = 0, \tilde{E} = 0$.

Equation (4.20) describes a wave of holes with a continuously eroding profile (fig. 4.4; 4.5). At moment $t = 1$, the wave front reaches the right boundary layer Π_ξ, the holes which entered this layer pass through it, and they begin to accumulate in the diffusion layer.

The field intensity $\tilde{E}_1 = \sqrt{1/(bt)}$ at the boundary of the regular layer when $\chi = 1$ continuously decreases in time, while the concentration increases. The complete solution in the entire interval between the two diffusion layers, at whose boundaries the zero values for the field intensity are accepted, is determined from the equations (4.20) and (4.13):

$$\tilde{E}_\Sigma = \tilde{E} + \bar{E} - \tilde{E}_1.$$

From system (4.20), when calculating the voltage drop on the regular layer, we get the following in dimensional form

$$U = \frac{2\mu_p j_+^2 t}{3(q\mu_n N_d)^2} + \frac{j_+}{qN_d\mu_n}(w - \frac{j_+ t}{qbN_d}) \text{ when } t < t_u;$$

$$U = \frac{2}{3}\sqrt{\frac{w^3 j_+}{q\mu_n\mu_p N_d t}} \text{ when } t > t_u . \qquad (4.21)$$

As was noted, the overall voltage on the p^+-n-n^+ structure is determined by the voltage on the regular layer. The solution to (4.20) makes it possible to find the total number of holes which passed through the regular layer and were accumulated in the diffusion layer by the n^+-- n junction, which is determined by the integral

$$P_{n^+} = \int_0^t j_p dt = \frac{2\mu_p j_+ t}{\mu_p+\mu_n}(1 - \sqrt{\frac{t_u}{t}}) + \\ + \frac{\mu_p^2 j_+ t}{\mu_n(\mu_p+\mu_n)}(1 - \frac{t_u}{t}) - N_d(\frac{j_+ t}{qN_d} - w), \qquad (4.22)$$

where $j_p = \mu_p pE$ is the density of the hole current at the boundary with the diffusion layer Π_ξ by the n^+-- n junction.

In reference [12], it was shown that expression (4.22) gives a very good description of the data of experiments when there are small τ_+ values and a small thickness of the diffusion layer (L << w).

In conclusion, the deep analogy between the problem we examined on the nonstationary double injection in the absence of recombination and the well known problem about stationary double injection with linear recombination should be emphasized. The formulation of the second problem is distinguished by the substitution of the derivative from the concentration with respect to time $\partial p/\partial t$ for the recombination term p/τ_p. The formal solutions also turn out to be very similar, and in the role of the current time t is the life time τ_p. Taking this remark into consideration, for example, the expression we obtained for the voltage (4.21) when $t > t_u$ is completely identical to the corresponding expression in reference [10].

In the very same reference, an attempt to take into consideration the influence of boundary diffusion layers in the stationary problem by means of meshing the solutions for the equivalent of the diffusion equation (4.16) with the regular solution of the equivalent of the diffusion equation (4.10), was also undertaken in this work. The equality of the concentration at the boundary between the diffusion and regular (drift) layers was accepted as the meshing condition. The meshing condition also determined the position of the boundary. In reality, the meshing condition should also have guaranteed the strict continuity of the rate of the minority carriers current j_p/j in order to eliminate the accumulation of carriers at the meshing boundary. Such an accumulation, which is permitted in reference [10], is physically absurd.

It is easy to show that, for a regular layer when there is a high injection level $j_p/j = 1/(1 + b)$, and for a diffusion layer, this condition is equivalent to the condition $\partial Q/\partial\chi = \partial p/\partial\chi = 0$. The exponential solutions for the diffusion layer in reference [10] satisfy this condition only when $\chi = \infty$, and not when there is a finite χ value. As was shown above, the meshing of a purely diffusion equation (4.16) with the equation for bipolar drift (4.10) is a mathematically incorrect operation. Therefore, attempts of this type are foredoomed to failure, which has happened in reference [10]. For the very same reason, the solution for (4.21) would have been incorrectly "adjusted" for those case when condition L >> w broke down, by means of narrowing the drift layer. The obtained solutions to equations (4.10) and (4.16) should be well fulfilled each in its own area: the first when $\chi > L$, and the second when $\chi < L$ (see fig. 4.3). In the area where they overlap ($\chi \approx L$), there exists an uncertainty zone where it is necessary to make a smooth transition from one solution to another. The size of

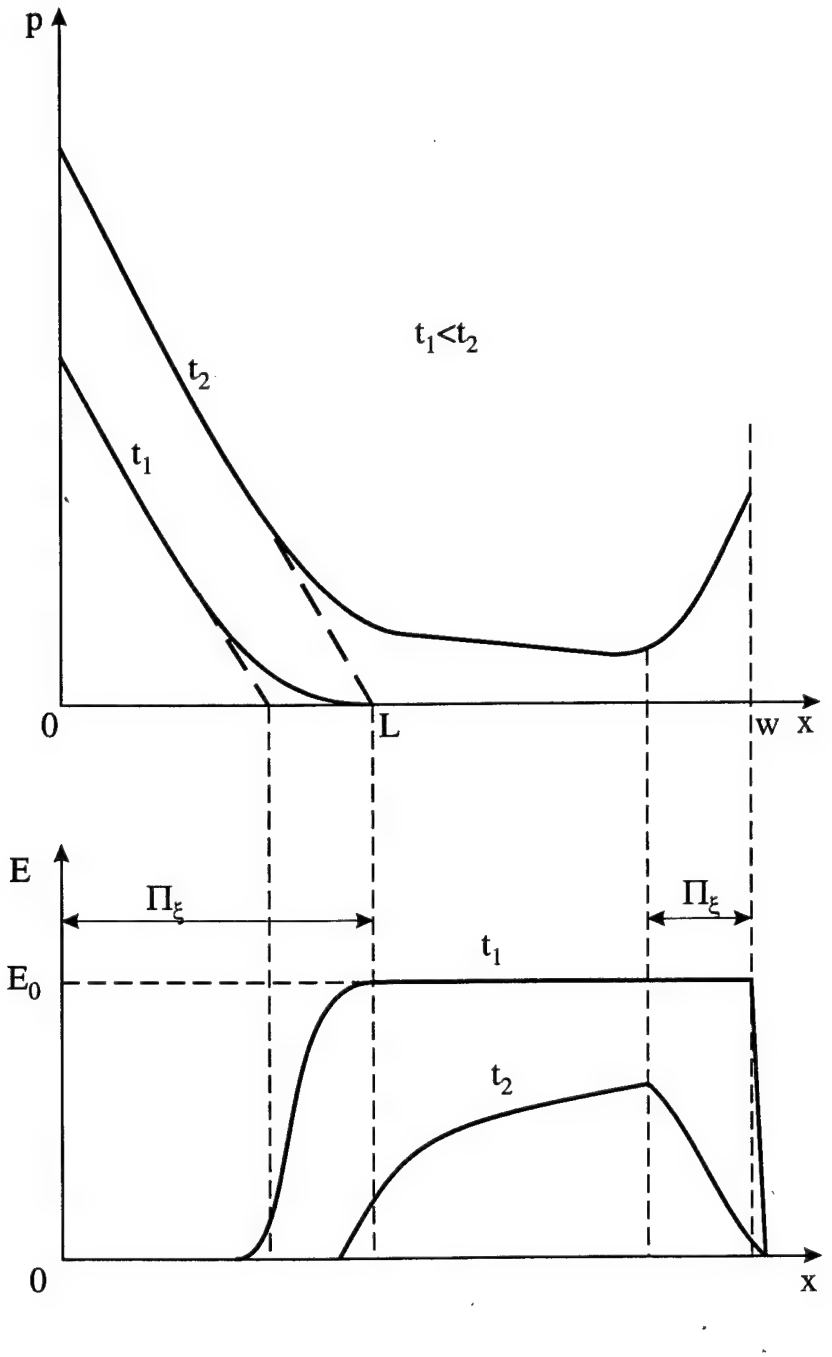


Fig.4.4

The distribution of the concentration of holes and the field intensity in the neutral n- region of the p^+nn^+ structure at the injection stage.

The dotted line is a linear approximation.

this zone can be evaluated with acceptable accuracy only from comparing it with a numerical solution, which have been done in our investigation and will be considered later.

However, since at the injection stage, as was shown above the diffusion layer has little effect on the complete voltage on the diode, the uncertainty in the overlap area, generally speaking, does not play a large role. As was shown below, at the plasma dispersion stage, the pattern changes sharply, and the effect of the diffusion layer can become very large.

4.2.2 Plasma extraction

It is obvious that the system of equations (4.7) and (4.8) also describes the plasma extraction process when there is a change in the sign of the current. That, which was previously stated about the small parameters and the appearance of boundary layers, remains in effect. However, the disposition and evolution of these layers can have a completely different character.

For further analysis, we will again specify a problem in agreement with the experiments in [12]. After a forward current pulse, with time length τ_+ and density j_+, passes through the p^+-- n -- n^+ structure, the current changes direction for the reverse, and during a time of approximately 10^{-9} s (which is small in comparison with the flight time through the n layer), a new constant value for the current density j_- is established. Thus, the initial state for the plasma extraction process by means of the back current corresponds to the state which is established when a straight current pulse with duration τ_+ passes through, and was examined in section 4.2.1.

During the extraction process, just like during the injection process, there exists a regular layer described by system (4.10); however, there will not be any points for the twisting of the curve of the potential χ_{00} and χ_{01}, and there will not be any barriers either. The carriers which had reached junctions p^+-- n and n^+-- n freely penetrate them, and at the same time the flow of carriers (of holes from the n^+ layer and electrons from the p^+ layer) in the n layer are absent. The existence of diffusion layers is now not obligatory as it was in the preceding case; however, the diffusion quasineutral layer Π_ξ which appeared at the injection stage, due to the relatively slow motion of the carriers within these diffusion layers, will also exist for some time at the extraction stage as well.

Let us examine the switching of the direction of the current after the stage of small duration injection ($\tau_+ < t_u$), when at this stage only one diffusion layer is formed, and this layer is positioned by the p^+-- n junction. After the current is switched, there begins the process of the extraction of the injected plasma, which has already been briefly considered at the beginning of this chapter. In this case, an SCR appears by the p^+--n junction, and the SCR expands and "eats up" the diffusion layer (fig. 4.2). The voltage drop on this region is added to the voltage drop on the regular layer, and, in a number of cases, the complete voltage on the diode can depend substantially on this. In order to determined the voltage drop on the SCR and the regular layer, let us examine for certainty the dynamics of the change in the concentration of the unequlibrium carriers in the diffusion layer and in the regular layer.

The diffusion layer.

As was already shown above, in the region by the p^+--n junction, in order to describe the concentration at the injection stage, a linear approximation of the form of (4.19) with a characteristic dimension $L \approx \sqrt{D\tau_+}$ (see fig. 4.4) is possible. It is natural to accept this distribution as the initial one for the following extraction stage, beginning at the moment the current is switched (t = 0). It is obvious that this initial distribution will continuously "erode" due

to diffusion. The eroded area is $\Delta \approx \sqrt{Dt}$ and the erosion may be disregarded when the condition $\Delta << L$, $t << \tau_+$ is fulfilled.

During plasma extraction, the current is carried through the p^+-- n junction only by holes. At the initial stage, this current is diffusion current, which gives the following slope for the back concentration gradient by the p^+-- n junction:

$$\frac{dp}{dx}_{x=0} = \frac{j_-}{2qD_p}$$

Just as in the injection stage, during extraction in the region near the p^+-- n junction, the distribution of the concentration is linear. For simplicity's sake, we will consider it to be linear throughout the entire region of the back gradient $0 < \chi < \chi_m$ (fig. 4.2). In this case, taking into consideration that the cross-hatched area is the complete number of carried out holes j_-t, using elementary geometric constructions, it is possible to obtain a law for the movement of point χ_m, which corresponds to the maximum concentration and which divides the concentration parameter into a toward gradient ($\chi > \chi_m$) region and a reverse gradient ($\chi < \chi_m$) region:

$$\chi_m^2 = \frac{4j_+Dt}{(j_+ + j_-)}$$

and it is also possible to find moment t_0 when the boundary concentration becomes equal to zero, and also the χ_{m0} value corresponding to this moment:

$$t_0 = \frac{\tau_+ j_+^2 b}{j_-(j_+ + j_-)(1+b)}; \qquad \chi_{m0} = L\frac{j_+}{j_+ + j_-}. \tag{4.23}$$

When there is a zero concentration at the p^+-- n junction, there appears a region of uncompensated space charge of the holes and ionized donors inside it.

In the general case, the velocity of motion of the SCR boundary ($\chi = a$) can differ from the velocity of motion of the concentration maximum ($\chi = \chi_m$). The first can be found in the following manner. From the condition for the retention of the flow of electrons in the coordinate system attached to the point of the maximum χ_m, we get

$$\frac{j}{2} = j_{n(\chi=\chi_m)} = qn_m v_m + \Delta j_n; \; \Delta j_n = \int_a^{\chi_m} n d\chi,$$

where $v_m = d\chi_m/dt$.

For the reverse gradient region ($a < \chi < \chi_m$)

$$\frac{dp}{d\chi} = \frac{j_-}{2qDp},$$

from here, from the geometric considerations (fig. 4.2), we get

$$\frac{j_-}{2} = \frac{qp_{m0}}{L}(1 + \frac{j_+}{j_-})(L - \chi_m)\frac{d\chi_m}{dt}. \tag{4.24}$$

When solving equation (4.24) with the condition

$$\chi_m = \chi_{m0} \quad when \; t = t_0$$

and taking into consideration that $\chi_m - a = 2qD_pP_m/j_-$, we find

$$a = L(1 - \sqrt{1 - \frac{2(t-t_n)}{\tau_+}(1 + \frac{j_-}{j_+})}). \tag{4.25}$$

The voltage drop on the SCR is determined in agreement with the Poisson equation using the total space charge $\rho = qN_d + j_-/v_p$. It is easy to show that when the condition $U_{SCR} > \varepsilon E_s/(2qN_d)$ or $a > \varepsilon E_s/(qN_d)$ is fulfilled, which, in its turn is correct for $N_d \approx 10^{14}cm^{-3}$, $U_{SCR} > 5V$ and $a < 10^{-3}$ cm, the field intensity in the SCR exceeds the E_s value for which the saturation of the carrier's drift velocity occurs. When $E > E_s$, we get

$$U_{SCR} = (\frac{qN_d + j_- / v_s}{2\varepsilon})a^2; \quad (4.26)$$

when $j_- < j_s = qN_d v_s$, it is possible to disregard the effect of the holes.

In the case of the fulfillment of the condition

$$U_{SCR} < \frac{\varepsilon j_-^2}{3\mu_p q^3 N_d^3} \text{ or } a < \frac{\varepsilon j_-}{2\mu_p q^2 N_d^2}$$

($U_{SCR} < 1$ V, $a < 10^{-4}$ cm when $N_d \approx 10^{14}$ cm^{-3}, $j_- = 10$ A/cm^2) the space charge of the holes drifting at a rate which is less than the saturation velocity plays the main role. Then it is possible to use the simpler expression .

$$U_{SCR} = \frac{2}{3}\sqrt{\frac{2j_-}{\varepsilon\mu_p}}\, a^{3/2}. \quad (4.27)$$

Regular layer.

In order to describe the processes in a regular layer bordering a diffusion layer, let us return to the differential equation for bipolar drift (4.10). It possesses an important property: **when changing the sign by the current and the time, the equation does not change**. This means that the processes described by them are symmetrical in time relative to the moment that the current is switched from forward to back when the type of boundary conditions is retained.

Strictly speaking, similar symmetry for the field processes is always fulfilled only for a diode with an infinitely thick base. In this case, any point of the concentration $p_{(\chi,t)}$ profile during extracting stage will return to the initial state along the very same characteristic along which it moved from the same state at the injection stage when there was any forward current pulse duration. In real devices, the profile points reached the n^+ -- n junction (boundary layer Π_η or $\Pi_{\eta\xi}$) , "diverge" from the characteristics of equation (4.10), and at the stage of the back current, symmetry should be broken down. This occurs if the duration of the forward current pulse is more than the flight time of the minority carriers through the base of the device that is , when $\tau_+ > t_u$.

When examining the restoration stage of a voltage rise, we will consider that the initial distribution of the electrons and holes created by a forward current fails to change during the current switching time then the concentration of the mobile charge carriers is described by the second expression of (4.20), from which it follows that the distribution of the electrons and holes in the regular layer is simply determined by the charge $j_+\tau_+$, which is injected by the forward current. Let us incorporate parameter $P_0 = qwN_d\mu_n/\mu_p$, which depends on the properties of the device only. When solving system (4.10), $E_{(x=0)} = 0$ was accepted for the boundary condition. This very boundary condition is also true for the back current stage. Then, when switching the forward current in the time interval $\tau_+ < t_u$, which corresponds to the injected charge $P < P_0$, taking into consideration expressions (4.21) and the symmetry noted above for the transition processes, the change in the reverse voltage in time has the following form:

$$U = \frac{2\mu_p(T-t)j_-^2}{3(q\mu_n N_d)^2} + \frac{j_-}{q\mu_n N_d}(w - \frac{\mu_p(T-t)j_-}{q\mu_n N_d}), \quad (4.28)$$

where $T = P_0/j_-$.

Expression (4.28) is true for $t < T$ (the read out of the time is conducted from the moment the forward current is switched to a back current). If the duration of the forward current pulse $\tau_+ > t_u$, which means the introduced charge $P > P_0$, then at the forward current stage the holes are accumulated by the n^+-- n junction, and when the current is switched from forward to back, they will have an effect on the field and the concentration by the boundary $\chi = w$,

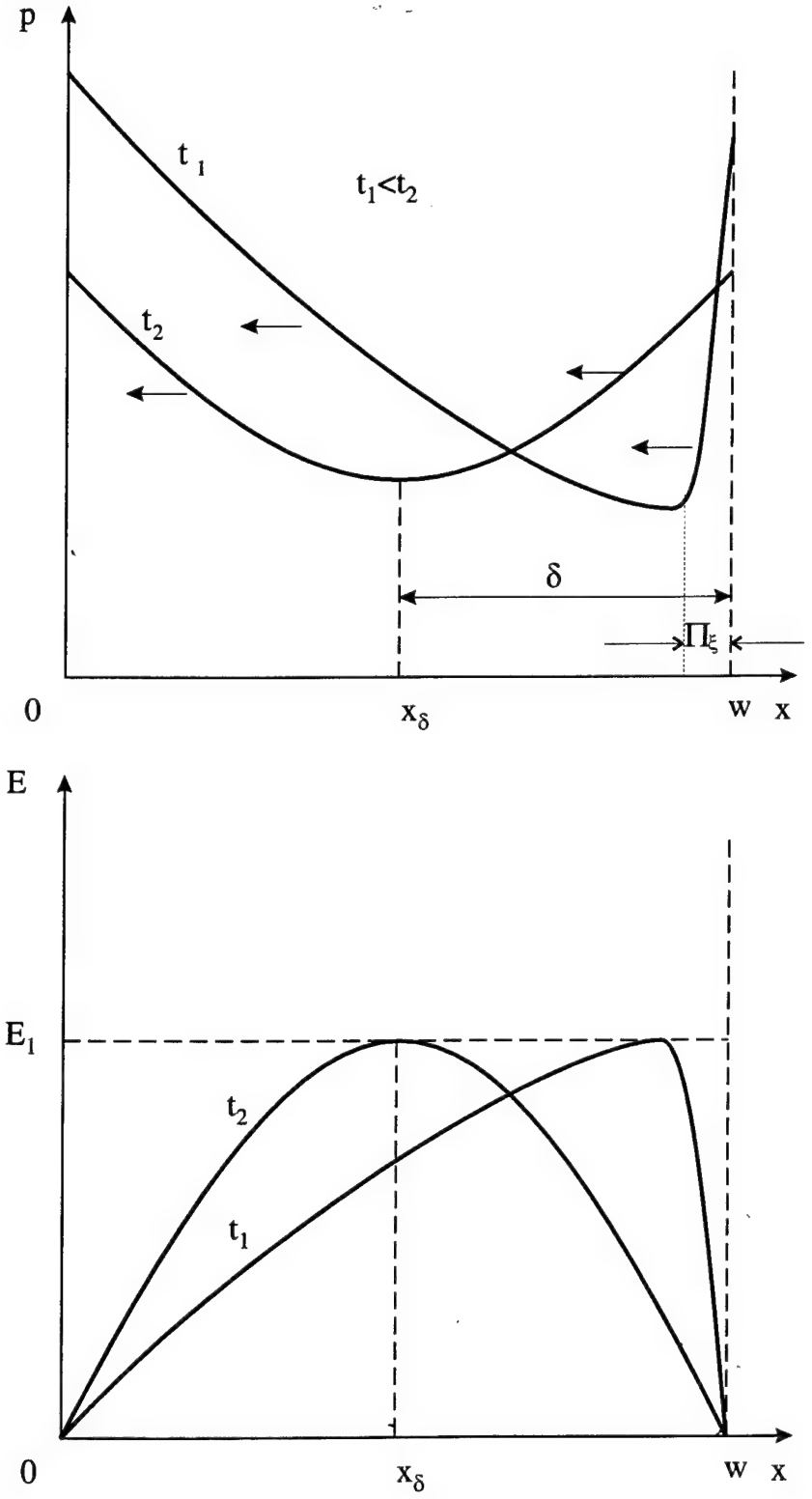


Fig.4.5
The distribution of the field intensity and concentration of the holes in a neutral n region when there are two diffusion layers in a p^+nn^+ structure.

which, as was noted above, leads to a breakdown in the symmetry of the transition processes. However, this is actually noticeable when $P > 3P_0$. Using the symmetry property of the forward and back transition process, we have the following for time $t < T - t_u$:

$$U = \frac{2}{3}\sqrt{\frac{w^3 j}{q\mu_n\mu_p N_d(T-t)}}, \tag{4.29}$$

where $t_u = wqN_d\mu_n/(\mu_p j_-)$ is the flight time when the current has a density of j_-.

In the time interval $T - t_u < t < T$ it is necessary to use equation (4.28).

In that case when the injected charge $P > 3P_0$, there is formed by the n^+ contact a noticeable plasma reservoir (diffusion layer Π_ξ), which at the stage of voltage restoration becomes a source of holes, and up to its exhaustion at the $\chi = w$ boundary the field intensity $E_w = 0$. Therefore, when the current is switched to the back direction, the diffusion layer by the $n - n^+$ junction begins to inject minority carriers into the regular layer, like the way in which the diffusion layer by the p^+-- n junction injected holes at the preceding stage of the forward current. Thus, from the diffusion layer by the n -- n^+ junction there is propagated in the depth of the regular layer a wave enriched by carriers (region δ in fig. 4.5), which is described by the already well-known drift equations (4.10) and their solution (4.20). Simultaneously with the expansion of region δ, there occurs the compression by the back current of the w-δ region, in which the points of the initial profile $P_{(x,t)}$, move along the very same characteristics of equation (4.10), as at the forward current stage. The boundary between the regions is determined by the expression

$$\chi_\delta = w - E_1\mu_p t,\ E_1 = \frac{j_-}{q\mu_n N_d}\sqrt{\frac{P_0}{P}}. \tag{4.30}$$

The voltage drop on the regular layer is:

$$U_p = \int_0^{\chi_\delta} E d\chi + \int_{\chi_\delta}^{w} E d\chi..$$

When using the boundary conditions accepted by us and equations (4.10) and (4.11), we get

$$U_p = \int_0^{\chi_\delta} E_1\left(\frac{\chi}{\chi_\delta}\right)^{1/2} d\chi + \int_{\chi_\delta}^{w} E_1\left(\frac{w-\chi}{w-\chi_\delta}\right)^{1/2} d\chi.$$

Integration yields

$$U_p = \frac{2}{3}E_1 w = const.$$

Thus, until the diffusion reservoir by the n^+-- n junction is exhausted, the voltage on the structure remains constant and is determined by expression

$$U_p = \frac{2}{3}j_-\sqrt{\frac{w^3}{q\mu_n\mu_p N_d P}}. \tag{4.31}$$

The moment of exhaustion of the diffusion reservoir can be determined after integrating the hole current at the stage of the extraction of the electron-hole plasma through the $\chi = w$ boundary. The charge obtained in this manner should equal charge P_{n+}, which was accumulated near the n^+-- n junction and was determined by expression (4.22). Taking into consideration that at the $\chi = w$ boundary, the high injection level condition is fulfilled, that is $p = n$, we get the ratio for the density of the hole and electron current $j_n/j_p = b$. Taking what has been stated into consideration, the moment of exhaustion of the diffusion reservoir is determined by the expression

$$t_{exh} = P_{n+}(1+b)/j_-$$

After the diffusion reservoir is exhausted, the holes begin to move away from the n^+- n junction, leaving behind them a region of uniform field in which $E_0 = j_-/(qN_d\mu_n)$. The velocity of expansion of this region is determined by the concentration of mobile charge carriers in the leaving front and by the current amplitude. This period of the junction process is represented

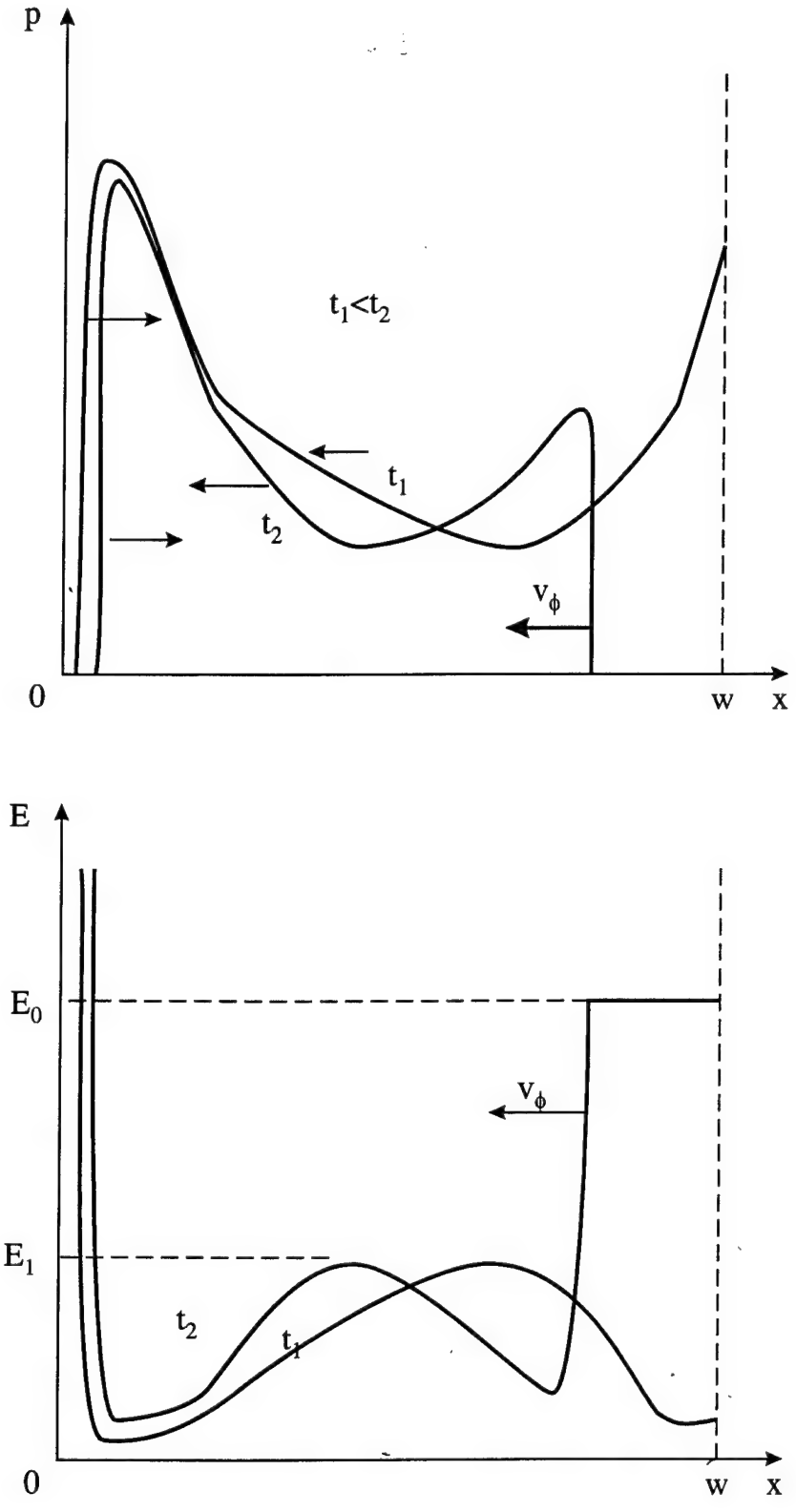


Fig.4.6
The formation of a running break during extraction of the plasma.

in fig. 4.6. The increase in the voltage from U_p to $U = E_0 w$ will correspond to it; at moment t_{exh}, a fracture in transient voltage curve should be observed.

In agreement with the noted property of equation (4.10), the points at which the concentration is less and the field intensity is more, will move more quickly; therefore, the slope of the jump E_0, which was formed near the $\chi = w$ boundary when the diffusion hole reservoir is exhausted, will continuously increase to infinity within the framework of equation (4.10); that is, a shock front is formed, which is a so-called break or shock front; physically its thickness is finite and is determined by diffusion.

The velocity of motion of the break v_φ depends on the maximum concentration of plasma in the break. In order to determine this concentration, it is necessary to solve system of equations (4.15), taking the diffusion into consideration, which is analytically impossible.

In order to evaluate the rate of growth of the voltage on the structure after the exhaustion of the diffusion reservoir near a n^+-- n junction, we will use a simplified analysis of this process. In front of the break, the holes will move more quickly than the break, that is, it will move away from it. This means that the concentration of holes in the break will decrease as it moves, and the speed will increase. During the time that the diffusion reservoir exists, charge $j_- t_{exh} = P_{n+}(1+b)$ will be removed from the base. In order that the back current bring out the remaining charge, time $t_1 = [P - P_{n+}(1+b)]/j_-$ is required. During time t_1, the break passes through the entire base w, and therefore the average rate of speed of the "break" is

$$v_\varphi = \frac{wj_-}{P - P_{n+}(1+b)}. \qquad (4.32)$$

After the formation of a break, the next regions will exist in the neutral part of n layer: the first, which is compressed, where the concentration of the electrons and holes is surplus (enriched region), and the field intensity E_1 is minimal, and the second, which is expanded, with a concentration of mobile charge carriers N_d and field intensity E_0. We will accept that the average field intensity in the first region during the entire time needed for pulling out the remaining charge after formation of the break is constant, $E_{av} = U_p/w$. Knowing the field intensity in both regions and the speed of movement of the boundary between them, we get the change in the voltage on the neutral region when $t > t_{exh}$

$$U = E_{av}[w - v_\varphi(t - t_{exh})] + E_o v_\varphi(t - t_{av}). \qquad (4.33)$$

It should be noted, that voltage drop on SCR in accordance with (4.25), (4.26) could significantly increase overall voltage on the structure.

After some moment enriched region disappears and space charge region meets the brake.

In conclusion, we would emphasize: during the process of the injection a fraction of the holes accumulated in the diffusion layer near a p^+-- n junction, when $p >> N_d$ equals b/(1+b) fraction of the complete number of holes in the n layer. The fraction of holes brought from the diffusion layer to the regular layer equals 1/(1+b). In the process of plasma extraction a fraction of the holes which had been pull out of the diffusion layer and which passed from the regular layer to the diffusion layer are the same as well (if recombination are disregarded). Therefore, the end of the process of pulling out holes from the regular layer after the passage of the concentration break through the layer exactly coincides with the moment of exhaustion, that is, of the disappearance of the remainder of the diffusion layer by the p^+-- n junction. The stated information signifies that the motion of the break is ended by the very sharp and complete depletion of the entire n layer from the nonequilibrium carriers. The further flow of the current through the neutral region is possible only due to the pulling out of equilibrium carriers -electrons with concentration N_d. The movement of electrons means the widening of SCR, the border of SCR moves with the speed of the electrons. Inside SCR the current is displacement current due to field intensity increase. In this case, the rate of growth

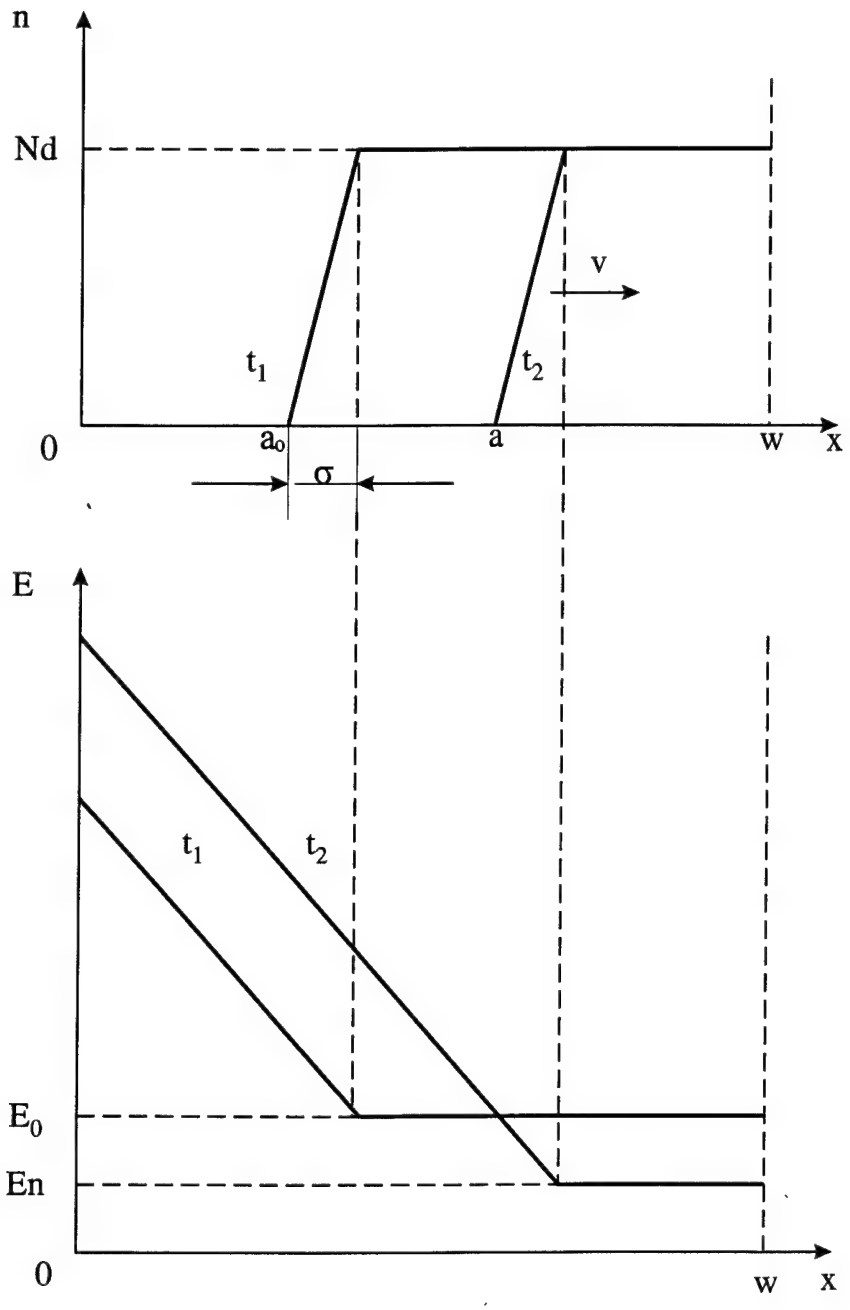


Fig.4.7
SCR restoration at the stage of the extraction of the equilibrium carriers

of the voltage sharply (by orders of magnitude) increases, even up to the maximum possible value [see formula (4.4)]

4.2.3 The restoration of the Space Charge Region Under Conditions of the Flight of Equilibrium Carrier

When the dimensions of the diffusion layer L are small up to the moment of the complete removal of the nonequilibrium charge carriers, the dimensions of the SCR and the voltage drop on it can be small ($L < 10^{-3}$ cm, $U_{SCR} \leq 10V$), while the voltage drop on the regular layer becomes a decisive factor.

The subsequent increase of the voltage on the SCR during the removal of the majority carriers under the conditions of a constant current flow will be limited either by the maximum voltage of the current source, or by the avalanche breakdown.

Let us examine the process of the restoration of a diode which is connected in a circuit in series with a purely active load R_l. Let us note that, due to the famous theorem about duality, this system can easily be converted in a system with a load connected in parallel to the diode, with the substitution of the voltage source for the current source.

The initial conditions, which are the distribution of the field intensity and the concentration of the mobile charge carriers in the base up to the moment of the complete removal of the minority carriers, are:

$$E_{(x,0)} = E_0 = \frac{j_-}{q\mu_n N_d};\ n_{(x,0)} = N_d;\ p = 0,\ a = a_0.$$

Let us accept this moment in time t = 0. Under the effect of the field, the electrons drift in the direction of the n^+ contact, leaving behind it the uncompensated space charge of the donors, due to which the field intensity (fig. 4.7) increases. The characteristic dimensions σ of the front of the wave of electrons leaving the base is equal to the Debye radius of the screening (less than 1μ in our case), and hereafter we will not take it into consideration. We will also accept the fact that the voltage source U_0 is constant. During the SCR expansion process, the external voltage is redistributed between the diode and the load resistance, the voltage on the diode increases and the voltage on the load drops. Due to this, the total current decreases, and the field intensity E_n, in which the electrons drift, also decreases. The processes ends when all of the external voltage is applied to the diode.

It is evident from the figure that the complete voltage on the diode

$$U = E_n w + \frac{qN_d a^2}{2\varepsilon} = U_0 - R_l S j_-,$$

where S is the area of the diode.

When differentiating this equation and substituting $da/dt = \mu_n E_n$ into it, after the replacement of the variable $\int_0^E E_n dt = \varphi$, we get in dimensionless form, taking expression (4.6) into consideration,

$$\beta\varphi_{tt} + \gamma\varphi_t + \varphi^2 - 1 = 0$$

$$j_- = E + \frac{1}{2\psi}E_t, \qquad (4.34)$$

where

$$\beta = \frac{\tau_c w}{\tau_M w_0 \psi^2};\ \gamma = \frac{w}{w_0};\ \tau_M = \frac{\varepsilon}{q\mu_n N_d};$$

$$\tau_c = \frac{R_H S\varepsilon}{w};\ \psi = 1 + \frac{\tau_c}{\tau_M};\ w_0 = \sqrt{\frac{2\varepsilon U_0}{qN_d}}.$$

The following values are accepted as single values: time $t_u = 2\psi\tau_M$; field intensity $E_u = U_0/(\psi w)$; current density ;$j_u = q\mu_n N_d U_0/(\varphi w)$. The j_- value is connected with E by equation (4.6).

The initial condition indicated above will take on the appearance $j_{(t=c)} = 1$; the second initial condition $\varphi_{(t=c)} = 0$.

Before solving nonlinear equation(4.34), we will evaluate the coefficients which are included in it. It is obvious that in the range of values for the parameters which interest us ($U_0 \approx 10^3$ V, $N_d \approx 10^{14}$ cm^{-3}, w $\approx 10^{-2}$ cm) we get $\gamma \approx 1$, $\tau_M \approx 10^{-10}$ s. The parameter $\tau_s = R_H C\ 10^{-9}$ s (C is the capacitance of the n region) and is an order of magnitude greater than the Maxwell constant. In this case, the $\beta \leq 0.1$ value turns out to be a small parameter in the case of the older derivative, which, as was previously noted, designates the appearance of a boundary layer (in the time dimension) with the fast changing of the variable.

In order to solve equations (4.34), we will use the method of asymptotic expansions with respect to small parameter b for singularly excited equations . We will transform system (4.34) into the following:

$$\beta E_t = -\gamma E - \varphi^2 + 1; \qquad (4.35)$$

$$\varphi_t = E.$$

We are searching for the solution to system (4.35) in the form of a series with respect to β with the following boundary functions:

$$E = E_0^* + \beta E_1^* + \Pi_0 E + \beta\Pi_1 E;$$

$$\varphi = \varphi_0^* + \beta\varphi_1^* + \Pi_0\varphi + \beta\Pi_1\varphi,$$

where the asterisk denotes the regular portion, and the Π denotes the boundary function.

When substituting the series into system (4.35) and when stretching the time scale $\tau = t/\beta$, we get a system of zero order equations with respect to β:

$$\begin{aligned} d\Pi_0 E/d\tau &= \gamma\Pi_0 E - (\Pi_0\varphi)^2 - 2\varphi_0^*\Pi_0\varphi; \\ 0 &= \gamma E_0^{*2} - \varphi_0^{*2} + 1; \\ d\varphi_0^*/d\tau &= E_0^*; \qquad (4.36) \\ d\Pi_0\varphi/d\tau &= 0. \end{aligned}$$

The solution of system (4.36), taking the initial conditions into consideration and the fact that $\Pi_0\varphi \to 0$ when $\tau \to \infty$ into consideration, has the form:

$$\Pi_0 E = Ae^{-\tau}; \qquad (4.37)$$

$$\varphi_0^* = \frac{e^{2/\gamma}-1}{e^{2/\gamma}+1}.$$

We get the final solution to system (4.34) when differentiating expression (4.37) and when substituting the result into equation (4.34) for the current density

$$j_- = \frac{4\exp(2t/\gamma)}{\gamma[\exp(2t/\gamma)+1]^2}\left\{1 - \frac{\exp(4t/\gamma)-1}{\psi\gamma[\exp(2t/\gamma)+1]^2}\right\} + \frac{\gamma-1}{\gamma}\exp(-\gamma t/\beta). \qquad (4.38)$$

The voltage on the diode is connected with the current by the Kirchgoff law:

$$U = 1 - j_- \frac{\eta}{1+\eta},\ \eta = \frac{\tau_c}{\tau_M}. \qquad (4.39)$$

As follows from expression (4.38), it has a quickly relaxing portion (the second member) with a characteristic time equaling

$$\frac{w}{w_0(\frac{1}{\tau_M}+\frac{1}{\tau_c})},$$

and a slowly relaxing portion (the first member) with time $(w/w_0)\cdot(\tau_M + \tau_c)$. The physical concept of the presence of two constant times consists of the following: when there is a great difference between τ_M and τ_c ($\tau_c >> \tau_M$, like in our case) at the initial stage, when the SCR is small and does not have an effect on the processes, the relaxation of the field is determined by the Maxwell time, and then, when the SCR covers the entire n layer, the diode is a pure capacity C, and the process of the current brake is determined by the relaxation of the circuit; that is, by the time constant $R_l C$.

Let us make note of the fact that in our case, $\psi\gamma >> 1$, it is possible to disregard the second factor in the squared brackets in expression (4.38) in comparison with the first.

The rate of growth of the voltage should increase with a decrease in R_H, since the load determines the initial current through the diode. This rate will be the maximum in the case of the initial current with density $j_s = qN_d v_s$, since in this case the electrons are removed from the p^+-- n junction with the maximum possible speed, which also provides for the fast growth of the voltage. The value for the rate of growth of the voltage U', obtained from expression (4.39), **will always be smaller than that previously obtained according to formulas (4.3) and (4.4) when the current density j_s is constant**.

The validity of the considerations made above have been checked by comparison of experimental data (transient currents and voltages) and calculated from formulas. The discrepancy was less than 20%.

4.2.4 Numerical modeling of the plasma injection and extraction processes

It was shown above that, for sufficiently large time intervals when the dimensions of a purely diffusion layer equals a noticeable fraction of the thickness of the n layer (L > 0.1 w), the effect of diffusion on the plasma extraction process becomes noticeable. Below we will examine the error of the analysis done above on the effect of diffusion in a "linear" approximation, after comparing this analysis with numerical modeling. We will immediately make note of the fact that a directly numerical solution for a singularly excited complete system (4.7) and (4.8) is very complicated. Such systems posses an exceptionally large computer - "cruelty " and unstability. Therefore, we used a simplifying system using the approach discussed above, after first making a series of additional explanations.

System (4.15), which describes the diffusion field transfer within the framework of quasineutrality can not describe the processes in an SCR which arises beyond the wave front. Up to moment ($t < t_0$) of the drop of the concentrations of electrons near the p^+--n junction to a value which is lower than the equilibrium, equation (4.15) are correct in the entire n layer, with the exception of thin (thickness of no more than 10^{-4} cm) boundary layers $\Pi_{\xi\eta}$ and Π_η by the p^+-- n and n -- n^+ junctions. As was already noted, the boundary conditions for the (4.15) system can be given either in the form of conditions on the ratios of the hole and electron flows, or by fixing the plasma concentration values at the boundary. When $t > t_0$, the region in which system (4.15) "works" also narrows, and the left boundary condition for the ratio of the current densities j_p/j must even be given on the running boundary of the SCR ($\chi = a$). In

the terms of the boundary layers, the expanding SCR corresponds to the Π_η layer. However, the equation for this layer will be different from equation (4.11), since the characteristic scale for the field intensity in the layer will no longer be close to one, and can be substantially higher. Therefore, during the derivation of the equation for layer Π_η, it is also necessary to deform the scale of the dependent variable E, as was done during the derivation of system (4.15). After doing the corresponding procedure for the search for the expansion, we arrive at the obvious result -- the Poisson equation, whose solution was already given previously [see expressions (4.26) and (4.27)].

Let us recall that between the SCR, that is the Π_η layer and the quasineutral layer where equation (4.15) is correct, there lies intermediate boundary layer $\Pi_{\xi\eta}$, where the simultaneous calculation of both the space charge and the diffusion is necessary. This means that it is not possible to correctly mesh the solutions for the SCR and the quasineutral region directly while omitting the $\Pi_{\xi\eta}$ layer. However, it is known that the solutions to the Poisson equation for currents limited by a space charge have little sensitivity to the condition at the weak field boundary ($\chi = a$), and easily come down to equations (4.26) and (4.27), in which the break of the concentration at the boundary ($n_{(x=a)} = \infty$) is allowed. In this case, a $\Pi_{\xi\eta}$ layer with a thickness of no more than 10^{-4} cm, the voltage drop on which is $kT/q \leq 0.1$ V, generally is not examined. Therefore, we will limit ourselves to a numerical solution of system (4.15), taking the running boundary $\chi = a$ into consideration, and for the SCR we will use the known analytical results for (4.26) and (4.27). Time derivatives do not enter into the equation for the SCR; that is, for the SCR, a quasistationary state is assumed, which physically signifies, as is true for the Π_η layer by the n^- region, the smallness of the flight time through the SCR as compared with the characteristic flight time of a carrier through the entire n region. Let us recall that in general, the ratios of the currents at the boundary, that is the boundary conditions in system (4.15), can not be given arbitrarily, but rather should be determined from the solution of the entire problem, while taking the parameters of the p^+ and n^+ regions and contacts into consideration.

It is obvious that this problem is unreliable not only due to the calculation difficulties, but also in view of the lack of sufficiently accurate data about the regions (the distribution of impurities, the presence of defects which yield active levels in the forbidden zone, etc.). Therefore, we used the boundary conditions for the currents as the problem's parameters.

The abbreviated system of equations for the entire interval $0 < \chi < w$, beyond the deduction of the boundary layers Π_η and $\Pi_{\xi\eta}$, that is, of the regions of the space charge, is obtained from equations (4.7) and (4.8) when values of parameters η and $\xi\eta$ strive toward zero. In this case, in equation (4.7) there remains the Dembers term which is proportional to Q_x. As was shown above, this term plays a part in neither the regular layer [see system (4.10)] nor the diffusion layer Π_ξ [see system (4.15)]; it is obvious that it can also be excluded in the simplified system ($n \to 0, \xi\eta \to 0$), which describes both layers. As a result, we get an equation which is completely equivalent to system (4.15), but is recorded in the undeformed scale. This result is obvious enough when taking into consideration the previously made notation about the fact that system (4.15) structurally includes system (4.10). We will make the transition in the obtained simplified system to a new variable, which is hole concentration p.

The abbreviated system of equations has the following form:

$$\partial p/\partial t = \partial j_p/\partial \chi;$$

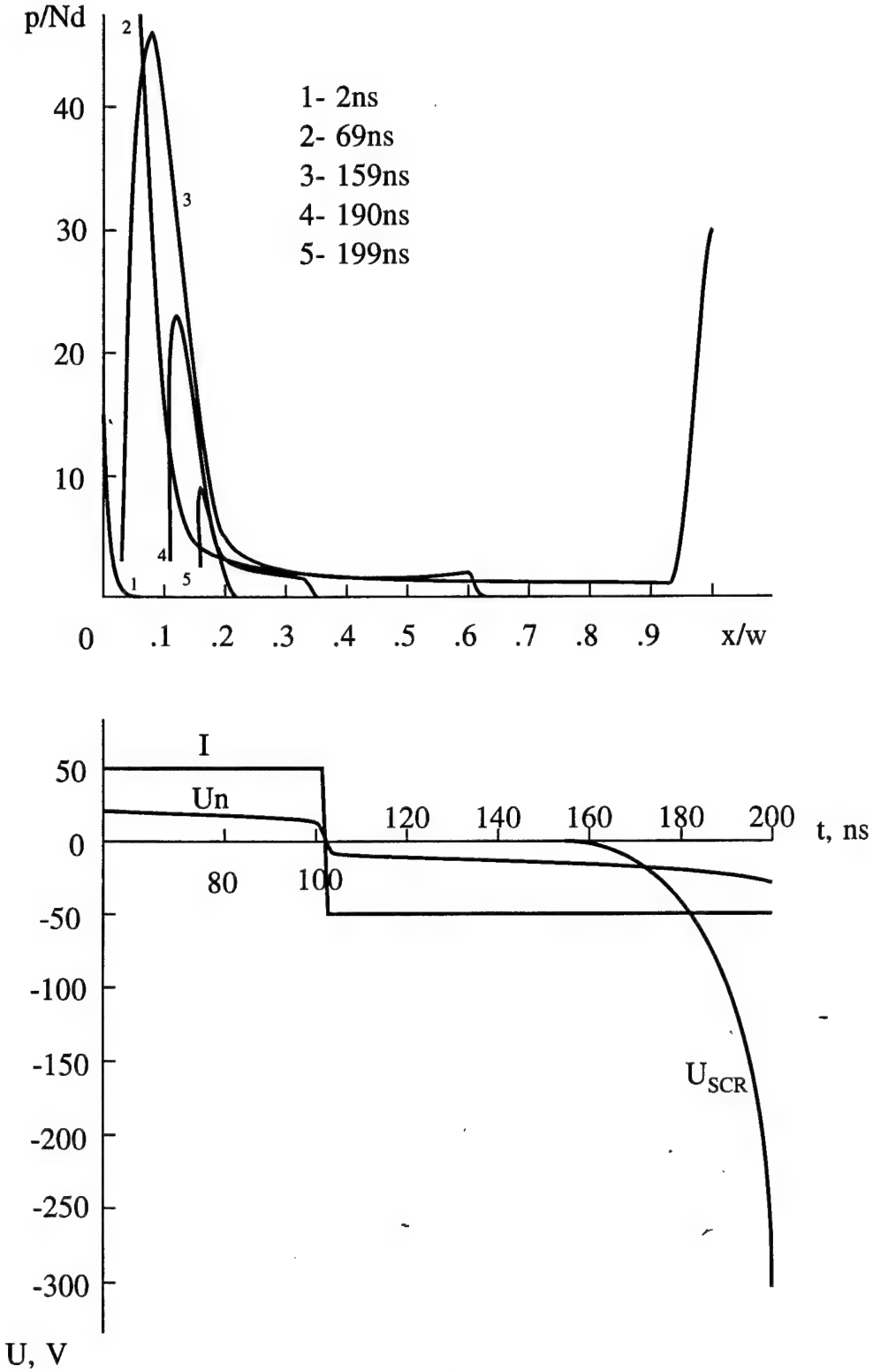


Fig.4.8
computer modelling of DSRD

$$j_p = Ap - B\partial p/\partial\chi; \qquad (4.40)$$

where

$$A = \frac{j}{p(1+b)+bN_d};\ B = \xi\left[\frac{(b-1)p}{p(b+1)+bN_d} + 1\right].$$

System (4.40) is identical to the well known simplified system used in calculations. Thus, the evaluation of this system, made using the methods of singular asymptotic expansions, can be seen as an mathematical substantiation.

System (4.40)(4.40) was numerically solved with the following initial and boundary conditions:

$$t = 0; \quad E = 0; \quad p = 0;$$

$$\chi = w; \quad j_p = j(1 - \gamma_n);$$

$$t < t_0; \quad \chi = 0; \quad j_p = j\gamma_p;$$

$$t > t_0; \quad \chi = a; \quad j_p = j\gamma_p,$$

where γ_p and γ_n are the injection coefficients for the p^+-- n and n^+-- n^- junctions.

System of equations (4.40) is reduced to a system of difference equations, and is solved by the known method using iterations.

The coordinate of the boundary ($\chi = a$) between the SCR and the neutral region at the plasma extraction stage is determined as the point in which the concentration of the holes, found from the solution of system (4.40), falls to a value lower than $N_d/2$. The value $P_{rp} = N_d/2$ is arbitrary (within the limits of an order of magnitude), however this, as the numerical calculations have shown, does not have an effect on the solution.

In fig. 4.8, there is shown the change in the distribution of the concentration of holes, depending on time. As is evident from the figure, the distribution qualitatively corresponds to the pattern described above. When a current pulse is supplied in the forward direction, a diffusion layer is formed near the p^+-- n junction, and from this layer the drift injection wave is propagated. After the wave passes up to the n^+ layer, a second diffusion layer is formed near the n^+--n junction. When switching the current to the back direction, back concentration gradient regions appear in the diffusion layers near the p^+-- n and n^+-- n junctions. After the drop of the hole concentration to zero, which corresponds to the drop of the electron concentration to the equilibrium value ($n_b \approx N_d$), an expanding SCR appears near the p^+-- n junction. We will recall that our calculation, that is, the solution of system (4.15), can not correctly describe the transfer of the carriers in the SCR.

When the concentration of the holes drops to zero near the n^+-- n junction, an equilibrium region is formed. As distinguished from the SCR, the neutrality condition is retained for the equilibrium region, since the flow of electrons from the enriched area compensates for the flow of electrons leaving for the n^+ region. We will note that near the p^+-- n^+ junction, an SCR arises due to the lack of the possibility of such compensation: the departure of electrons from the p^+-- n junction is not compensated by the arrival of electrons from the p^+ region in the case of the given condition $\gamma_p = 1$; that is, $j_n = 0$ when $\chi = 0$. Thus, from fig. 4.8 the origin of two exhausted of the holes regions (near the p^+-- n and n -- n^+ transitions) with sharply shaped boundaries (shock fronts) moving to meet one another is

evident. The increase in the voltage on the quasineutral region is mainly determined by the expansion of the exhaustion region near the n -- n^+ junction. However, the rate of growth of the voltage on the exhaustion region is relatively small, and the voltage drop can not exceed the drop determined by the initial resistance of the material of the n region at the given current. In fig. 4.8 there is given the curve for the voltage drop on the SCR, calculated according to expressions (4.26) and (4.27) with the use of values for the boundary of the SCR $\chi = a$, taken from fig. 4.8. It is obvious that the voltage on the SCR makes a substantial contribution to the complete voltage drop on the diode, beginning from the middle of the extraction process, and is a decisive factor in the last approximately 20 ns before the complete exhaustion of the entire volume. In other words, before the super-fast restoration stage (carrying out the equilibrium carriers), which takes up 1-2 ns, there exists a stage for the increase in the voltage with a duration of approximately 20 ns, smoothing out the sharpness of the transition to the super-fast stage.

We will compare the results of the analyses given above with the results of the numerical calculation. From what has been stated above, it follows that the distribution of the holes for the forward current stage can be characterized by the following values:

1. Near the p^+-- n junction [see formula (4.19)], by means of the dimensions of the diffusion layer

$$L_0 = \sqrt{\frac{4b}{1+b} D_p \tau_+}$$

and the maximum concentration $p_{mo} = j_+ L_0/(2qD_p)$.

2. Near the n^+-- n junction, by means of the corresponding L_w and p_{mw} values, which are determined according to the very same formulas. However, the readout of the time here is conducted from moment t_u, which corresponds to the end of the flight of the holes through the n region ($t_u = j_+ w/(qN_d b)$, and all of the n and p indices change places.

3. In the regular layer, by means of the smallest concentration value, which, based on system (4.20), equals

$$p_{\min} = \frac{N_d}{1+b}\sqrt{1+\frac{j_+ b\tau_+}{qwN_d}}. \qquad (4.41)$$

In table 4.1 there are given the corresponding parameter values which were calculated according to the given formulas (numerator) and those obtained based on fig. 4.8 (denominator). It is obvious from the table that the analytical calculation yields somewhat increased values for p_{m0} and L. This is connected with the fact that part of the injected charge turns out to be concentrated in the transitional, unconsidered meshing region between the purely diffusion and regular layers.

The correspondence of the results of the analyses and of the numerical calculation at the pumping stage (forward current) is completely satisfactory (the difference is less than 20%).

Table 4.1

Moment in time,ns	$p_{mo}*10^{16}$ cm^{-3}	L_0 , μm	p_{mw} x 10^{15} cm^{-3}	L_w , μm	$p_{min}x10^{13}$ cm^{-3}
20	1.0/0.8	8.58/8.25	--	--	--
60	1.8/1.4	15/18	4.8/2.0	10/6.25	6.9/4.5
100	2.4/1.9	20/25	7.2/3.7	16/10	18.0/9.1

For the diffusion layer near the n^+-- n junction, a simplified analysis gives a p_w value which is almost two times greater than numerical modeling, the discrepancy is connected

with the fact that in the case of the derivation of the formulas which determine the concentration in the diffusion layers, a high injection level at which $j_p/j_n = 1/b$ is supposed. For fig. 4.8, this condition was not strictly fulfilled ($p_{min} \approx 10^{14}$ cm^{-3} = N_d). Therefore, the fraction of the hole current is really less than the accepted one, and the accumulated hole charge is correspondingly less as well. When increasing a complete charge which has passed through the diode, the condition for the high injection level is fulfilled more strictly, and the discrepancy in the data from the analysis and the modeling decreases (when τ_+ is increased to 500 ns, the discrepancy is less than 30%).

The large concentration value p_{min} at the boundary of the regular layer and the second diffusion layer, which was obtained analytically, is also connected with the initial presumption about the large level of injection at the boundary from the start. At the initial stage, the condition $p > N_d$ is not fulfilled, and the flow of holes from the regular layer to the diffusion layer is smaller. With an increase in the charge which passed through the structure, the accuracy of the calculation of p_{min} increases.

For the extraction stage, moment t_0 of the equation of the concentration to zero near the p^+-- n junction and the position of the concentration maximum χ_{m0} at this moment are important. The data from the elementary theory and the numerical calculation correspond well: in agreement with formulas (4.23), χ_{m0} = 9.6 μm ; t_0 = 37 ns; in agreement with fig. 4.8 these values were 10 μm and 36 ns.

The values for the thickness of the SCR in μm are compared below:

	139 ns	154 ns	169 ns	182 ns	185 ns
Formula (4.25)	0.5	3.67	7.7	13.3	15.2
Fig. 2.13	0.7	5.5	13.2	22.5	27.5

It follows from this that the results of the numerical modeling yield a larger value for the dimensions of the SCR. This is connected with the fact that, as early as for the moment of the beginning of the formation of the SCR ($t = t_0$) , point χ_{m0} from which the back gradient region is formed, approaches the boundary of linear approximation, or the meshing region, where this approximation begins to break down.

Naturally, subsequently the back gradient region goes deeper into the meshing region. Thus, the meshing region has a weak effect on the total appearance of the distribution of the carriers and on the voltage drop in the case the forward current, but it can have a very strong effect on the calculation of the thickness of the SCR and the voltage drop on it during extraction -- in that case when the back gradient region ends up the meshing region.

As follows from fig. 4.8 at the extraction stage, the distribution of the holes in the diffusion layer in the forward gradient region, that is, when $\chi > \chi_{m0}$, changes very weakly even in the case of the equality of the forward and back current ($j_+/j_- = 1$), when the duration of the extraction and pumping stages are identical. At first glance, the further expansion of the region during time τ_- for value $L_- = \sqrt{D\tau_-}$ should have been expected. The explanation of the effect of the "freezing" of the distribution consists of the following: in the case of the forward current and a high injection level, the speed of the holes $v_p = j_p/pq$ by the p^+-- n junction is determined by both as a diffusion current (j_d) and as a drift current (j_E), since these components are equal and are directed in one direction, which then doubles the complete current as compared with the diffusion current. In the case of a rapid change in the direction of the current (in the case of the equality of absolute values of the current density), the gradient and the diffusion component of the current retain the very same value ($j_d = j/2$). Now, the drift component ($j_E = j_d = j/2$) changes direction for the opposite one, so that the complete current of the holes due to such compensation decreases, and the distribution of the holes is "frozen". The

complete electrical current in this forward gradient region is carried by electrons. Due to this factor, the linear distribution of the concentration, used when deriving formula (4.25), **with an L_0 value which is fixed for the extraction stage, turns out to be a good enough approximation for the case $j_+/j_- \approx 1$ as well.**

One should also pay attention (see fig. 4.8) to the extremely sharp collision (or smacking) of the SCR and the exhaustion region before the moment of the complete disappearance of the unequlibrium carriers, lasting 1-3 ns, and not resolvable in a numerical experiment. Such a sharpness is caused by steepness of the fronts of the regions which are smacking together. The slope of the fronts is determined by the diffusion, and equals $dp/d\chi \approx j/(qD)$, from which for the width of the front we get $\delta \approx p_m qD/j$ where p_m is the concentration at the front. Taking into consideration that, at the moment of the smacking together, the concentration at the front in the regular layer equals p_{min}, according to formula (4.41) it is possible to evaluate the δ value and the smacking time

$$\tau_\delta = \frac{\delta}{v_\delta} \approx (\frac{qp_{min}}{j})^2 D \approx \tau_+ \frac{j+qN_d D}{j^2 w_n}, \qquad (4.42)$$

where $v_\delta = j/(qp_{min})$ is the speed of the front.

The evaluation of (4.42) gives $\tau_\delta \approx 0.5$ns; that is, a value lying beyond the threshold of the step for the derivation of the mesh which is used for constructing graphs 4.8 . It is precisely this smallness of the interval which provides the sharpness of the transition to the super-fast restoration stage when there is the condition of a preceding sufficiently slow increase in the voltage on the SCR.

The comparison of the analytical solution with the numerical modeling given above has shown that the analytical approach with the linear approximation of the diffusion layer yields deliberately lowered values for the voltage on the arisen SCR. The error begins to catastrophically grow as the SCR boundary grows closer to the diffusion layer. Thus, for this case, it is necessary to use only a numerical solution.

As was already shown (see fig. 4.8) for $\tau_+ \approx 100$ ns, the voltage on the SCR at the 180th nanosecond equals 63 V, and it exceeds the voltage drop at the exhaustion region. By the 195th nanosecond, U_{SCR} reaches almost 300 V, creating a noticeable voltage "precursor" in front of the super-fast restoration section. When τ_+ increases up to500ns and more, slow precursor increases sharply up to 800 volts.

Such a sharp increase in the precursor when increasing the duration of the pumping of the diode are clearly seen in experiments and strongly worsens the switching properties of the diodes.

4.2.5 The Effect of the Accumulation of Carriers in the Emitter Layer

The mechanism described above for the formation of an SCR near the p^+-- n junction is connected with a deficit of electrons which are carried by the current from the junction, and which compensates for the space charge of the holes and the impurity. The process of formation of the SCR can be delayed if, from the direction of the p^+-- n junction, a flow of electrons is introduced which is equal to the flow of electrons leaving the diffusion layer for the regular layer. If this flow is continuous, then the SCR is not formed, and the restoration of the voltage does not begin at the p^+-- n junction even after the exhaustion of the entire n layer. Therefore, the flow of electrons must be interrupted before the moment that the p^+-- n junction is reached by the front of the exhaustion region (shock front).

It is possible to realize such an additional flow of electrons to the p -- n junction by different methods. For simplicity's sake, we will examine a symmetrical p^+-- p -- n -- n^+

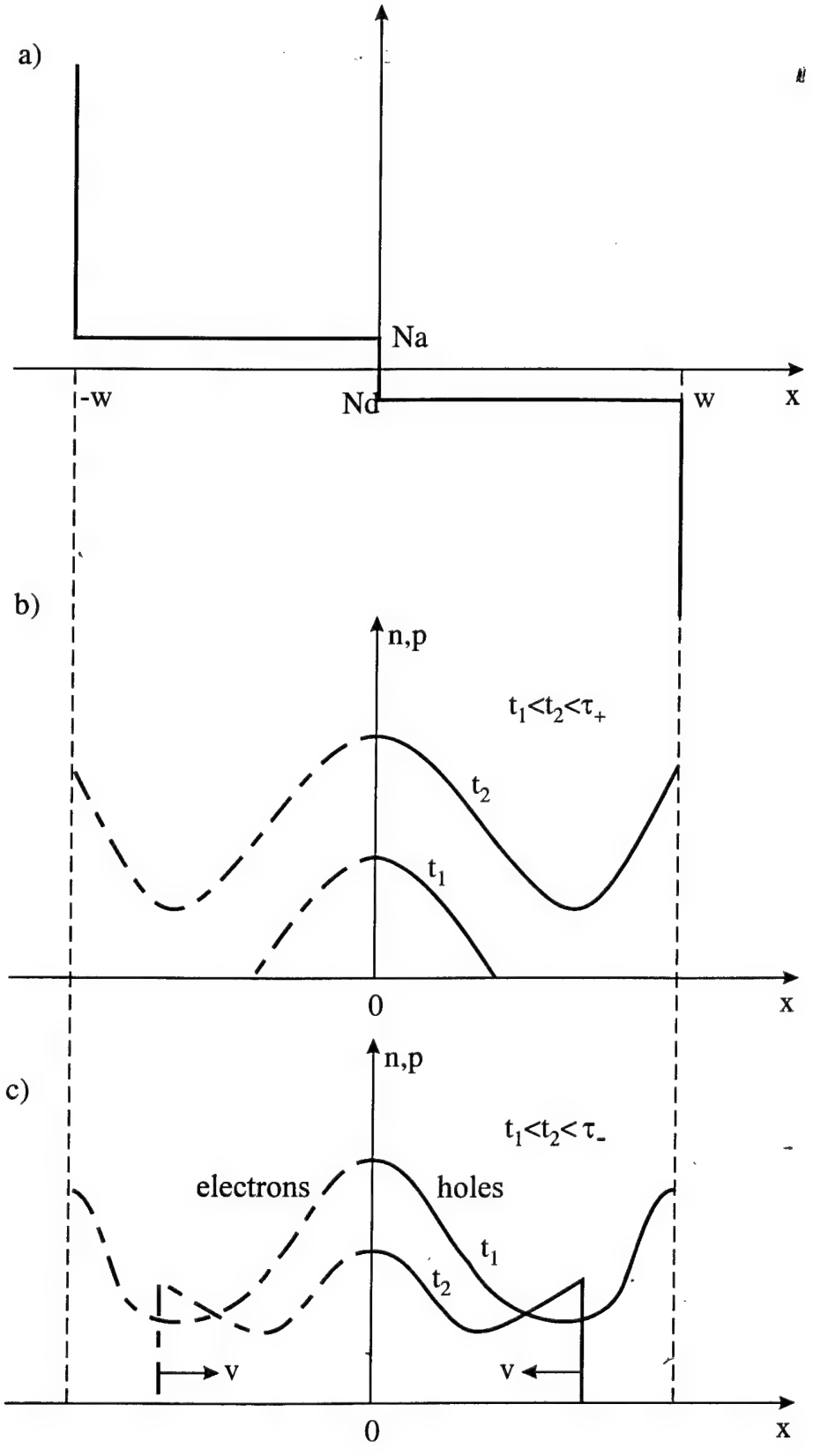


Fig.4.9
The distribution of the concentration in an ideally symmetrical p^+pnn^+ structure when the current is switched.

structure in which the difference between the p and n regions consists only of the substitution of the acceptors for donors (fig. 4.9), while the electrons and the holes differ only by the charge sign; at the boundary of the p -- n junction, the injection coefficients for the holes in the n region and for the electrons in the p region equal 1/2.

Let us examine the processes for the transfer of the holes only in the n region; for the p region the transfer of electrons will be completely analogous.

Let a step change in a forward current with density j_+ (fig. 4.9) be given for the structure. Just like in the p^+-- n -- n^+ structure, the holes injected from the p region to the n region will begin to drift in a field with intensity $E_0 = j_+/(q\mu_n N_d)$. The composition of the current during drift is determined by the known ratio $\gamma_{pE} = j_p/j = p/(p + N_d)$. The density of the hole current at boundary $j_p = j_+/2$ is greater than in the regular layer ($j_{pE} < j_+/2$); this means that near the n -- p junction holes will be accumulated. Just like in the p^+--n -- n^+ structure, near the p -- n structure the p_m value will increase. As is known from the solution of the symmetrical problem for a stationary case with equilibrium contacts, the concentration $p_m \approx j^{1/2}$ can also significantly exceed the concentration of the impurity: $p_m >> N_d$ [8].

It should be expected that, during the transition process, at some moment t_m, the condition $p_m > N_d$ will begin to be fulfilled. From this moment, the distribution of the carriers in the regular layer will correspond to law (4.20), while near the n -- n^+ junction, a diffusion layer Π_ξ is formed. The difference between the p^+-- n -- n^+ structure and the symmetrical p^+-- p -- n -- n^+ structure will consist only of the fact that in the latter the complete number of holes accumulated in the diffusion layer near the p -- n junction will be a great deal less.

After changing the direction of the current (fig. 4.9), in the regular layer there begins a redistribution of the concentration which is completely analogous to that which occurred in the p^+-- n -- n^+ diode. After the exhaustion of the diffusion reservoir near the n^+-- n junction, a running exhaustion wave front arises, behind which the equilibrium carrier region remains. As follows from equations (4.5), in an ideally symmetrical structure, condition $j_p = j_n$, when there is a high injection level ($p \approx n$) can be fulfilled only if $dp/d\chi = dn/d\chi = 0$; that is, if a back gradient region does not arise near the p -- n junction (as distinguished from the p^+-- n junction).

At the extraction stage, a flow of electrons with density $j_n = j/2$, neutralizing the space charge of the holes and donors, continuously goes from the p region to the n region; that is, an SCR is not formed. At the moment of the exhaustion wave front's approach to the p -- n junction and the interruption of the flow of the holes through it, the flow of electrons is also interrupted simultaneously due to the passage of a similar front in the p region. The duration of the process of the smacking of two exhaustion waves at the p -- n junction is determined by the very same processes as the p^+-- n -- n^+ structure [see formula(4.42)] and is just as small. The subsequent flow of the current will be accompanied by the departure of the main carriers, and by forming and the rapid expansion of the SCR in both directions from the p -- n junction, similar to the process examined in point 4.2.3. It is obvious that, in the case of ideal symmetry, the fast restoration of the voltage is possible after as long a current flow as is wanted (the lag of the fast restoration is determined by accumulated charge and reverse current j_-). Let us note that when there is the equality of the injection coefficients in the p^+-- p and n -- n^+ junctions to one of charge loss in the p^+-- p -- n -- n^+ structure will be determined only by recombination.

The ideal symmetry of the structure is unrealized, and the real pattern can differ markedly from the described one...

Let us qualitatively examine the effect of the factors which breakdown symmetry.

1. The mobility of the electrons and holes in silicon differs by $b = \mu_n/\mu_p = 2.8$ times. In the case of the symmetrical distribution of the impurity in the p and n regions, the ratio of the charges accumulated in the n region ($p_n = j_+\frac{\tau_+}{1+b}$) and in the p region ($N_p \approx j_+\frac{b\tau_+}{1+b}$) at the pumping stage equals $P_n/N_p = 1/b$. Precisely such an equation is also correct for the charges pulled out as well; therefore, the moments of the exhaustion of regions p and n will coincide, and the fast restoration of the voltage will begin in these regions simultaneously. The rate of expansion of the SCR [$v = (j_-/qN_A d)$] in them will also be identical. The rather small difference in the rates will rise in the case of currents with a density of j_- corresponding to the saturated velocity v_s. This difference only slightly slows down the rate of growth of the voltage with respect to the "ideal" rate, which is equal to double the rate of growth of the voltage on an asymmetrical structure (p^+-- n -- n^+) where the SCR expands only to one side.

2. An increase in the concentration of the doping impurity in the p region leads to an increase in the hole injection coefficient γ_p . In the threshold case ($N_A >> N_d$), a transition to an asymmetrical structure occurs.

3. The most substantial factor in the breakdown of symmetry is the loss of the charge during injection. In asymmetrical diodes, the charge losses due to recombination or penetration to the rear contact (n^+ region) leads only to a decrease in the effectiveness of the pumping, while having no effect on the pattern of the transitional process (only small rate of the voltage increase at the delay stage, and a small decrease of the voltage rise rate after the delay stage.

In symmetrical diodes the fraction of the charge accumulated in the diffusion layer of n the region near the p -- n junction decreases, with respect to the entire charge accumulated in the n region. Therefore, in the case of charge losses in the p region (for example, due to a sharp decrease in the lifetime of minority carriers during the process of creating a p -- n junction), the charge accumulated in the p region is extracted earlier than in the n region. The period of time, during which the current of electrons through the p -- n junction exists, is shortened, and in the n region the SCR begins to expand near the junction. Due to the small concentration of carriers in the diffusion layer, the rate of growth of the voltage on the SCR will be a lot larger than when $\gamma_p = 1$. On the transition characteristic of the device (at the delay stage), a "precursor " will appear which worsens its switching properties. Charge losses are also possible in the n region for example, due to the penetration of the carriers to the rear contact (n^+ region). If in this case the complete charge in the p region turns out to be less than in the n region, then the SCR will begin to be restored first in the p region. Just as in the preceding case, due to the lowered concentration of the carriers in the diffusion layer, the fast growth of the voltage at the delay stage begins, increasing the "precursor".

The qualitative approach stated above have been compared to numerical modeling. As was the case in the section 4.2.3, only one n layer has been considered. Charge accumulation in the p region was simulated by choose of the injection coefficient of the p^+-- n junction $\gamma_p = 1/(1+b) = 0.26$. The modeling shows that the fraction of the charge accumulated in the $\Pi\xi$ layer by the p -- n junction, when $\gamma_p = 0.26$, decreases sharply, almost by an order of magnitude as compared to the case of $\gamma_p = 1$. At the very same time, at the extraction stage, the SCR arises 15 ns later than when $\gamma_p = 1$, since after the forming of the SCR, its rate of expansion is approximately three times less, and the rate of growth of the voltage decreases by

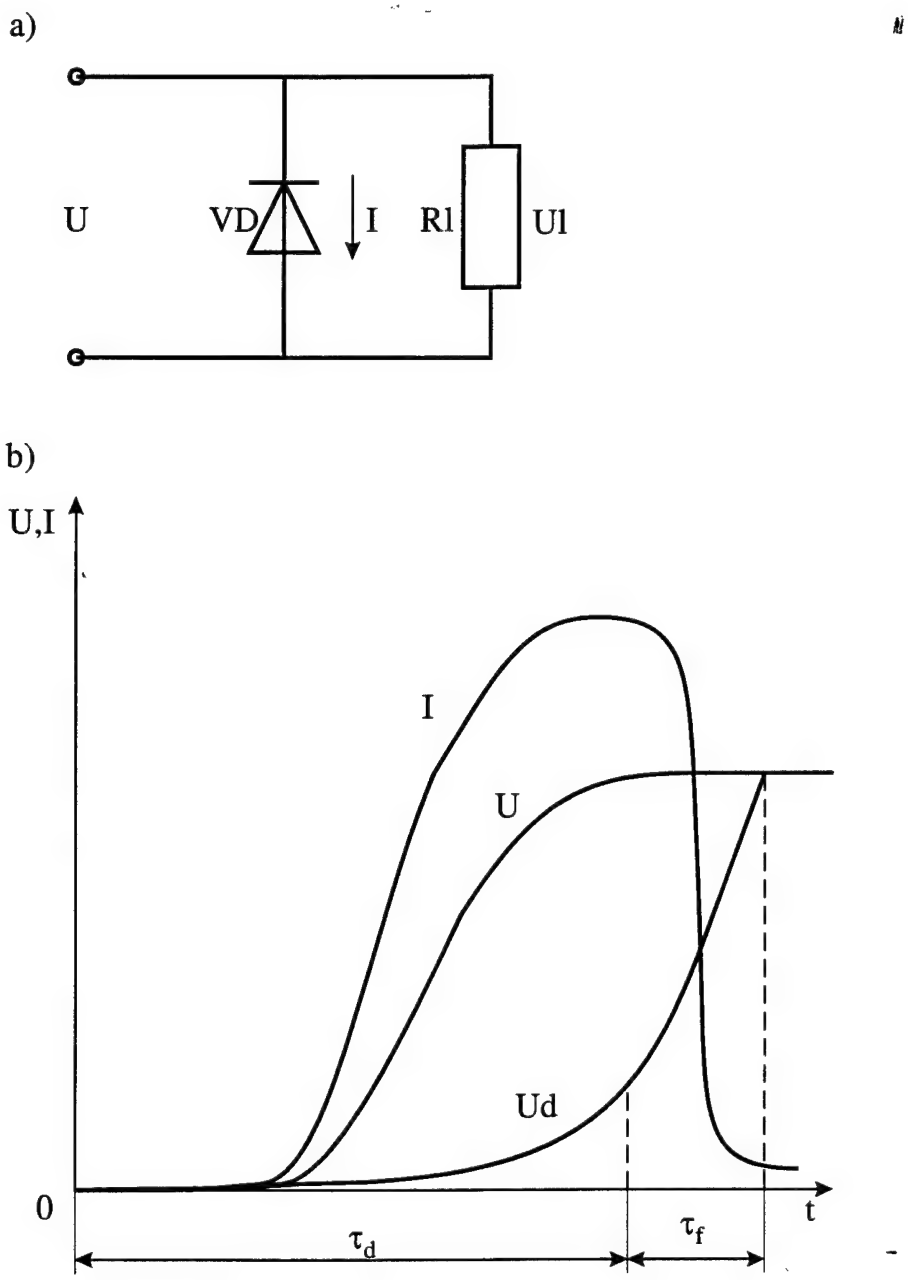


Fig.4.10
Schematic diagram of the switching on and operating principle of a drift diode with sharp restoration.

almost an order of magnitude. Thus, when $\gamma_p = 0.26$, the calculation does not exclude the fact of SCR formation itself, in spite of the conclusion made above based on general qualitative concepts. Such a divergence is caused by the following factor. During the numerical computation, the γ_p coefficient was given as constant for the entire time of the process. At the boundary (between the diffusion layer near the p -- n junction and the regular layer), the composition of the current changes with a change in the concentration $\gamma_p = j_p/j = p/[(1+b)p + bN_d]$ from the value $1/(1+b)$ when $p >> N_d$ to $p/(bN_d)$ when $p << N_d$. Correspondingly, the fraction of the electron current is $j_n/j = 1 - \gamma_p$ when the decrease in the concentration increases in comparison with the value $1 - \gamma_p = 0.74$ when $p >> N_d$. Therefore, when decreasing the concentration in the diffusion layer near the p^+-- n junction to a value $p \approx N_d$, the departure of electrons to the regular layer exceeds their arrival from the p region and the SCR arises.

The duration of the stage of the existence of the SCR before the complete exhaustion of the volume equals $\tau_{SCR} \approx \tau_- N_d w/(2P)$, where $P = \int_{\tau_+} jdt$ is the complete accumulated charge in the diode. **Thus, in a symmetrical p^+-- p -- n -- n^+ structure when there is an increase in the accumulated charge, the influence of the SCR on the extraction process decreases.**

4.3 Diode like switches with one p-n junction

The above considered physics of electron -hole plasma injection and extraction in p^+nn^+ structures is the physical basis of new devices -Drift Step Recovery Diodes (DSRD). Actual devices structures and their use in circuits can differ from those used in above consideration. In this section we will consider: how these differences influence the device performance, what changes or additions must be made for practically used DSRD, what improvements are possible .

Just as it is true for the well known step recovery diodes (SRD), when switching a short forward current pulse to a reverse current, the DSRD over a certain amount of time τ_d (delays of the beginning of fast restoration) possesses a high conductivity, and is capable of allowing a significant current when there is a small applied voltage. The DSRD, switched in parallel to the load, shunts it during the entire time τ_d (fig. 4.10). Then the conductivity of the DDRV drops sharply, and the entire current is switched to the load. Thus, the DSRD operates like an opening switch.

As distinguished from the SRD which permits operation with a direct current in the forward direction (we call it the pump current), the DDRV operates, as a rule, in the pulse mode. The parameters of the pumping , the amplitude, and the duration are very tightly connected with the efficiency of operation of the DDRV as a switch.

The discussion given above on the physics of nonstationary processes in diodes makes it possible to select the optimum ratio for the values of the DDRV's design elements and its operating mode.

4.3.1 Current form influence.

Analysis of the injection and restoration processes was done for a special (step) forms of the voltages and/or currents, which simplifies the explanation for the physical pattern of the processes considerably. However, for practical purposes, the case of a smooth change, for

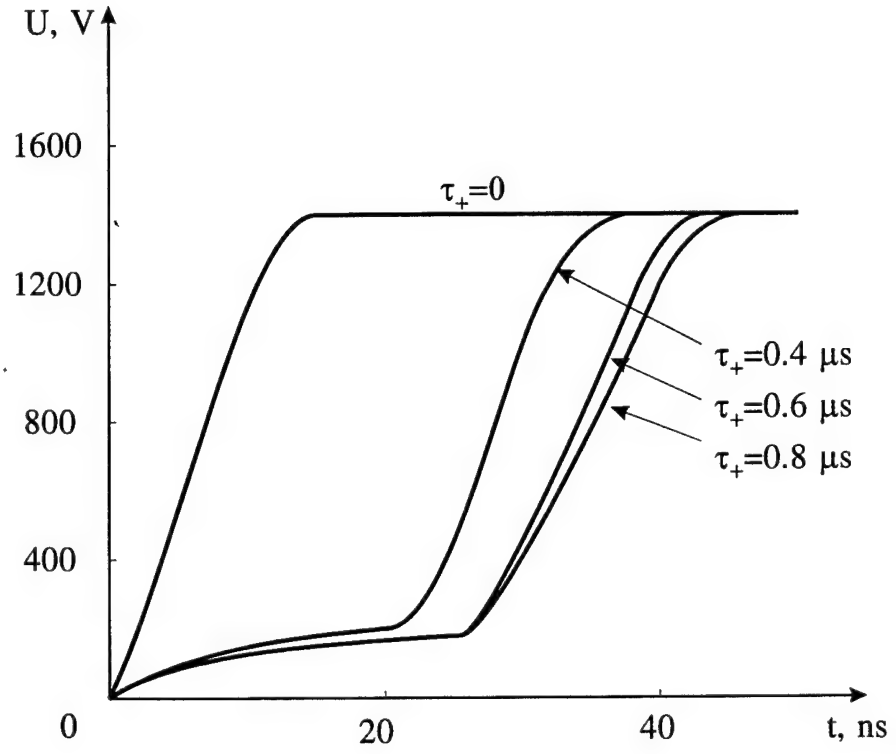


Fig.4.11
The restoration of the voltage on the p+nn+ structure with a small lifetime τ_p= 0.5μs for different forward current durations.

example sinusoidal or linear, in the currents given by a circuit which is external to the diode is interesting.

It is clear that qualitatively, the pattern of the processes has been retained. At the very same time, the operation of the diode in pulse circuits up to the moment of fast restoration is characterized by parameters which are integral with respect to time. Such parameters are: the complete charge, accumulated in the diode, of the unequlibrium carriers $P=\int_{\tau_+} Jdt$; the fraction of the complete charge which remained in the diffusion layer, determined by the ratio of the mobility; the dimensions of the diffusion layer , determined by the duration of the time interval $L=\sqrt{D\tau_+}$. It is obvious that these parameters have little sensitivity or are completely insensitive to a specific form of current pulses. The fast restoration processes in conditions of the pulling out of equilibrium carriers is determined only by the value of the current from which this process begins.

The experimental curves confirms the conclusion which could be made from (4.20) about the identical distribution of the electrons and holes when there was an equal injected charge. The deviation from this rule when τ_+ is large more than 400 ns is explained by the fact that, due to the increase in the dimensions of the diffusion layer by the p^+-- n junction, the SCR dimensions also increase; therefore, the voltage drop on the SCR becomes a decisive factor.

The extraction process is characterized by the distribution of the plasma injected at the first "pumping" stage. In structures for whom the life time was decreased to τ_p =0.5 μs by means of gamma radiation, when increasing the duration of forward current to values $\tau_+ > \tau_p$, the plasma extraction process is determined only by the lifetime τ_p . The remark made previously about the analogy between low duration τ_+ nonstationary processes and stationary processes with a small lifetime τ_p is confirmed by this [see fig 4.11]. This feature is of practical significance. It makes possible to use constant current pumping of DSRD.

4.3.2 Efficiency.

In practical work, a voltage pulse with the minimum possible front duration τ_f, when there is the minimum voltage "precursor" before the front, must be obtained in the load, since the "precursor" substantially lowers the efficiency of the device.

As was shown, at the end of the extraction of the electron/hole plasma by means of the backward current a distribution is formed which corresponds to the equilibrium (unmodulated) state of the diode; that is, in the entire base $n = N_d$ and $p = 0$; in this case the density of the current through the device $j = qN_d\mu_n E_0$. Condition (4.4) for the maximum rate of growth of the back voltage when the main charge carriers are flying out of the base requires that the restoration current density $j_s \approx qN_d v_s$, while $E_0 = E_s$. When increasing the density a sharp increase in the "precursor" begins, and when decreasing it, the front lengthens. For the given amplitude of the shaped voltage pulse, the optimal current density can be established by means of changing the area of the device, or the load resistance. In this case, the amplitude of the voltage on the neutral region just before the front $U_{nu} \geq wE_0$. In order to obtain the minimum "precursor", it is necessary to manufacture a DSRD with the minimum possible base thickness,. In reality the thickness value should correspond to the width of the SCR in the case of the voltage of a stationary avalanche breakdown of the p -- n junction.

As was noted above, when increasing the duration of the pumping, the influence of the diffusion layer near the p^+-- n junction begins to have an effect, and it is precisely here that a wide SCR arises earlier than all the holes leave the base. Thus, due to the voltage drop

on the quasineutral region U_{nu}, an additional voltage drop U_{SCR} arises near the p^+-- n junction. Consequently, in this case the "precursor" is determined by the sum $U_{SCR}+ U_{nu}$.

Let us examine the effect of U_{SCR} on the operation of the device in more detail. The characteristic thickness of the diffusion layer $L = \sqrt{D(\tau_+ + \tau_-)}$ is the parameter which determines the value of U_{SCR}. This condition for the smallness of U_{SCR} immediately imposes a limitation on the total time of the pumping and delay $(\tau_+ + \tau_-)$.

In addition, as was mentioned previously, the U_{SCR} value depends significantly on the character of the distribution of the impurity in region L. When there is partial compensation by the acceptor impurity of the n layer donors, which occurs in the process of creating the p -- n junction, the space charge of the doping impurity (totaled with the hole charge) can be decreased. Double diffusion (for example of Boron and Aluminum), which makes it possible to create the required compensation area on the "tail" of the rapidly diffusing impurity, is the optimum. However, as follows from expression (4.27), when $j_- \approx j_s$ and $L \approx 3 \cdot 10^{-3}$ cm even in the case of complete compensation, the U_{SCR} value levels with U_n, which imposes a natural limitation on L($L << 10^{-2}$ cm) and on the total time ($\tau_+ + \tau_- \leq 5 \cdot 10^{-7}$ s) in kilovolt range diodes.

A substantial decrease in U_{SCR} is achieved in "quasisymmetrical" p^+-- p -- n -- n^+ diodes with charge accumulation in the emitter p layer. However, in devices manufactured by diffusion technology, in the p and n layers there is an impurity concentration gradient N_a-N_d. For this reason, the value of the parameter $j_s = qv_s(N_a - N_d)$ changes along the n-layer, and it becomes impossible to retain the optimality condition for the fast restoration stage ($j_- = j_s$). As a consequence, such structures when providing a large permissible value $\tau_+ + \tau_-$ have a somewhat smaller maximum rate of growth of the voltage as compared with the p^+-- n -- n^+ structures.

The principle for the formation of fast restoration of the voltage which was discussed above can also be applied to low voltage structures (100 V and less): then the duration of the forward current pulses and the delay time should be a lot less than in high voltage structures. In this case, the effectiveness of the device, when determining the ratio of the complete time $\tau_+ + \tau_-$ to the duration of the restoration time τ_f [$\kappa = (\tau_+ + \tau_-)/\tau_f$] worsens (K determines compression ratio in time compressing circuits). Actually, $\kappa \approx v_s L^2/(Dw)$; that is, $\tau_+ + \tau_3 \approx L^2/D$, and to get the required sequence for the extraction processes (at the delay stage, this is the extraction of the unequlibrium carriers, and at the restoration stage this is the extraction of the equilibrium carriers), the condition L/w < 5 is necessary. Consequently, the effectiveness decreases with a decrease in L as a square.

It is possible to decrease the effect of the duration of the pumping when using a DSRD with a small lifetime of the minority carriers .. In section 4.3, the analogy between stationary processes with small carrier lifetime and nonstationary fast processes has already been given. This analogy is expressed in the equivalency of the pattern for the switching processes in diodes with very small minority carrier lifetimes ($\tau_p \approx 0.5$ μm) in the case of a large duration for the forward current pulses ($\tau_+ >> \tau_p$) and in diodes with a large lifetime ($\tau_p > 10$ μs), with a small duration for the forward current pulses $\tau_+ = 0.5$ μs (fig. 4.11). However, the complication of the technology for manufacturing the devices and the sharp increase in the losses for pumping is not justified by the certain systemic advantage of a DDRV with a small lifetime (simplification of the decoupling of the main current circuit and the auxiliary circuit of the pumping current) in the majority of variants for their use which are interesting for practical purposes.

4.3.3 Synchronization.

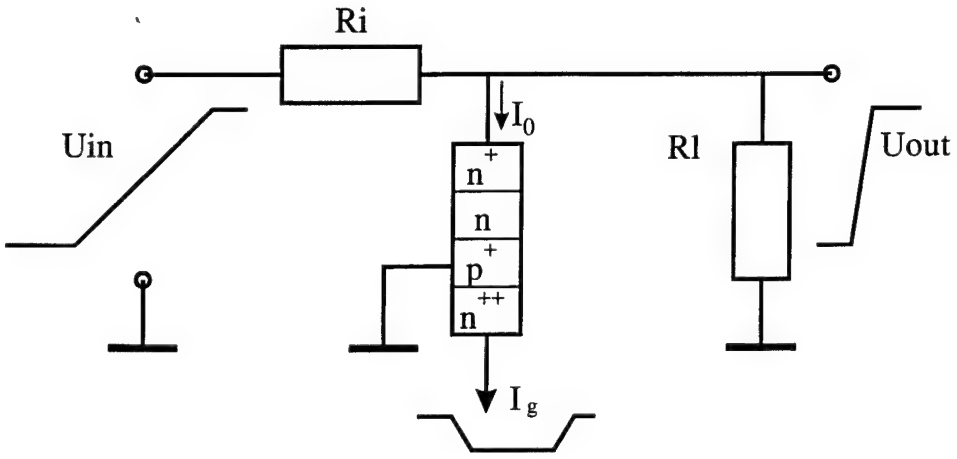


Fig.4.12
Drift step recovery transistor

Those devices in which charge losses are absent, that is, $\tau_p >> \tau_+ + \tau_3$, possess a number of advantages. First of all, the stability of the moment of switching for them is determined only by the stability of the circuits which shape the forward and back current pulse, and it is not connected with the device's parameters. (It follows from the experimental data that a temporary instability is less than 50 ps and is determined only by the stability of the synchronization circuits and the power sources). In addition, in the case of the connection in series in the circuit of several DDRV for increasing the switching voltage, the total current for pumping and extraction automatically guarantees the synchronicity of the moment of the voltage restoration on all of the devices. However, when shaping pulses whose front τ_f has a nanosecond duration, the synchronization should also be with an accuracy in the ones and fractions of nanoseconds.

In the case of the complete duration of the cycle for the flow of the current, $\tau_+ + \tau_- \approx 1$ μs, the relative permissible charge loss should be $\tau_f/(\tau_+ + \tau_-) \leq 10^{-3}$, that is, it is extremely small. Even in the case of purely recombination losses and a lifetime of 100 μs, the τ_p value should be controlled with an error of no more than 10%. When decreasing τ_p, the required accuracy grows correspondingly. Thus, during the operation of a DSRD assembled in "stacks", due to the synchronicity requirement, it is necessary to decrease the duration of the pumping and the τ_+ value as compared with those permissible for a single device.

As was shown, at the stage of the fast growth of the voltage, the rate of voltage rise is simply determined by the degree of doping of the n-layer and the density of the flowing current. Therefore, the difference in the voltage on the diodes is determined by the difference in their operating area and the degree of doping. Both of the latter parameters determine the volt-farad characteristic of the device as well. As practical experience has shown, the preliminary selection of diodes according to a dispersion of the volt-farad characteristics within the limits of 20% makes it possible to connect in series dozens and more diodes without a noticeable total widening (elongation) of the front.

It follows from the operating principle of the DSRD that internal positive current feedback (current regeneration) is lacking in it, and the entire switching cycle is equal to the flight time of the carriers through the n-layer of the structure. This determines the main advantage of the fast restoration in the case of switching using DSRDs. In addition, due to the lack of regeneration, there is no localization effect (filamentation of current), connected with the development of the instability of the uniform distribution of the current, which is typical for such switching devices as thyristors, avalanche transistors, etc. This makes it possible, using a simple increase in the working area of the device, to increase the switched current practically up to the thresholds determined only by the circuits parameters or by skinning (Part 1, (2.34)).

4.4 Transistor like switches with two p-n junctions.

Fast step recovery processes in p-n junction may be incorporated in more complicated structures with two (transistors - DSRT) or even with three (thyristors-DSRTh) p-n junctions.

Let us consider three layers transistor like n^{++}-- p^+-- n -- n structure, shown in fig. 3.2. In this structure p^+-- n -- n^+ part is similar to one of DSRD considered earlier. This device is connected as shown in fig. 4.12 (with the base grounded). Initially no voltage on collector from external source is applied. High current gating pulse with density j_g is applied in forward direction to base emitter n^{++}-- p^+ junction during τ_+ time. After diffusion time $\tau_d \approx W^2p/2Dn$ electrons from emitter reach the collector p^+-- n junction and p^+-- n junction is forced into saturated state. After this moment the process of holes injection from p+ layer to n-layer

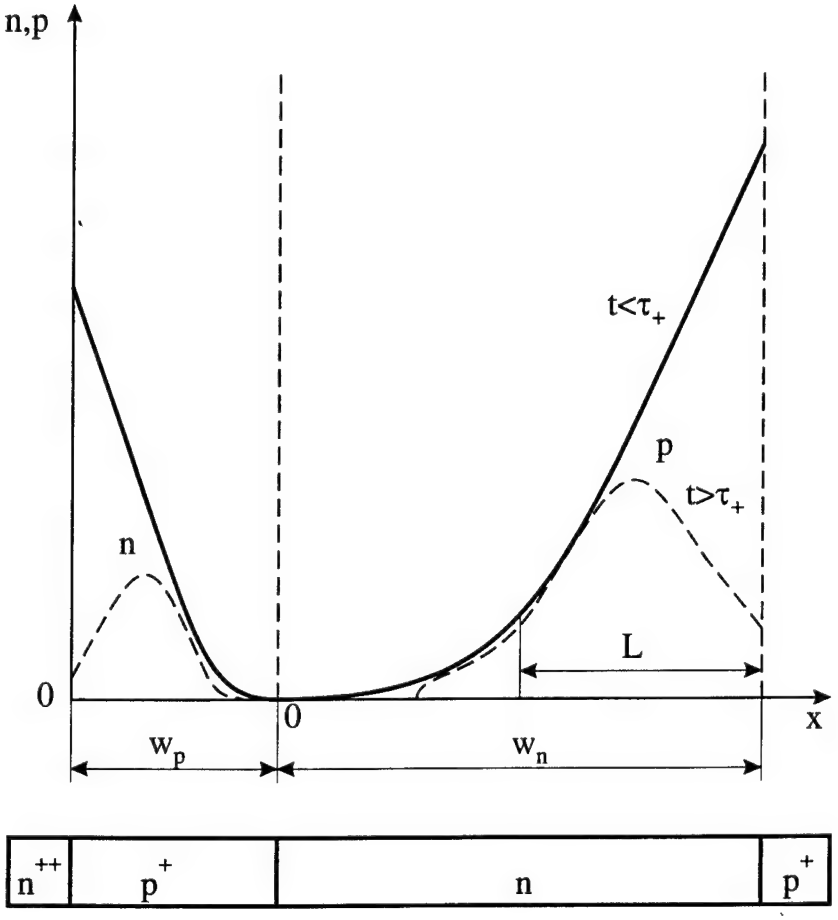


Fig.4.13
Fast restoration in four layers $n^{++}p^{+}np^{+}$ structure.
(carriers concentration).

begins. The transfer of holes through the n-layer is diffusion in nature due to the absence of voltage applied, therefore the width of the enriched by carriers layer appeared in n-layer (L) is

$$L \approx \sqrt{D(\tau_+ - \tau_d)} \text{ or } L \approx \sqrt{D\tau_+}, \quad (\text{if } \tau_d < \tau_+)$$

The well know Fletcher condition for high injection level links the carriers density on both sides of p^+-- n junction:

$$p_p n_p \approx N_d n_p \approx p_n^2 \approx n_n^2 \tag{4.43}$$

If the transistor amplification gain is large ($\beta >> 1$), the carrier distribution in p+ base is linear. Assuming that the carrier distribution in n-layer is near to linear, as well we can get:

$$n_p W + \frac{j_+ W_p^2}{2qD_n} + \frac{(n_n - N_d)L}{2} = j_+ \tau_q \tag{4.44}$$

The first and second terms in (4.44) are total number of electrons in p^+ base, the third is the number in n-layer. By use of (4.43), (4.44) can be transformed into

$$\frac{p_n L}{2}\left(\frac{2p_n W_p}{N_d L} + 1\right) + \frac{j_+}{q}(\tau_+ - \tau_d) \tag{4.45}$$

Equation (4.45) shows that the most part of carriers are accumulated in the diffusion region , having L-thickness, of n layer when condition $\tau_+ >> \tau_p$ is fulfilled.

When the gate pulse is turned off, the collector voltage U_0 (fig. 4.12) is applied and collector current due to accumulated charge starts. Enriched by free carrier region of n layer ($\chi < L$) (fig. 3.2) is in high conducting state and its voltage drop is small. The remained part of n layer is in equilibrium state and the voltage drop (U_{eq}) is $U_{eq} = j_c(W_n - L)/q\mu_n N_d$, where j_c collector's current density, limited by external circuit $j_c = U_0/SR_i$ (A - collector area, R_i - external resistor).

The plasma dispersion process in transistor is similar to the process in DSRD (reverse concentration gradient, SCR restoration). The main difference is that no drift nonlinear wave appears due to the absence of holes outside enriched diffusion region. Electrons, coming from p^+ base into n layer, help to suppress SCR restoration and to improve switching properties. To get this improving it is necessary that all electrons have been removed from the p^+ base before the enriched diffusion region disappears completely. If not, the rate of the collector current brake is slow and is determined by a slow process of electron movement into the n layer across p^+ base.

The second main (additional to $\tau_+ >> \tau_p$) condition for fast current brake and voltage restoration for transistors is the same as for DSRD: enriched diffusion region of n layer (L) must be shorter than SCR width (W_{SCR}) ($L << W_{SCR}$). The limit to voltage rise rate should be determined by (4.38) as for DSRD. Actually, in DSRT the collector current concentrates on border between base and emitter electrodes. The maximum current density in DSRT is limited by the same as for DSRD condition $j_c < qv_s N_d$. That is, total collector current must be smaller than is possible for the DSRD of the same area. Lower average current leads to somewhat (1.5 ÷ 3 times) lower voltage rise rate.

It should be noted, that the described mode of operation is possible, when collector voltage source U_{in} supplies constant voltage and collector current limited by the resistor j_{cm} is smaller than gating current ($j_{cm} << j_g$).

4.5 Thyristor like switches with three p-n junction

It is interesting to evaluate the possibility of super-fast restoration in four layers structures (thyristor-like) with more complicated interplay of electrons and holes currents.

Let us consider thyristor-like structure n^{++}-p^+-n-p^+ (fig.4.13), having n-layer width (W_n) much more than the width of p^+-layer (W_p) and connected into the circuit fig.4.14. Initially the capacitor C is charged in "forward" for the thyristor polarity. After triggering of the thyristor (gating pulse I_g is applied), the capacitor discharges via the thyristor and inductor L.

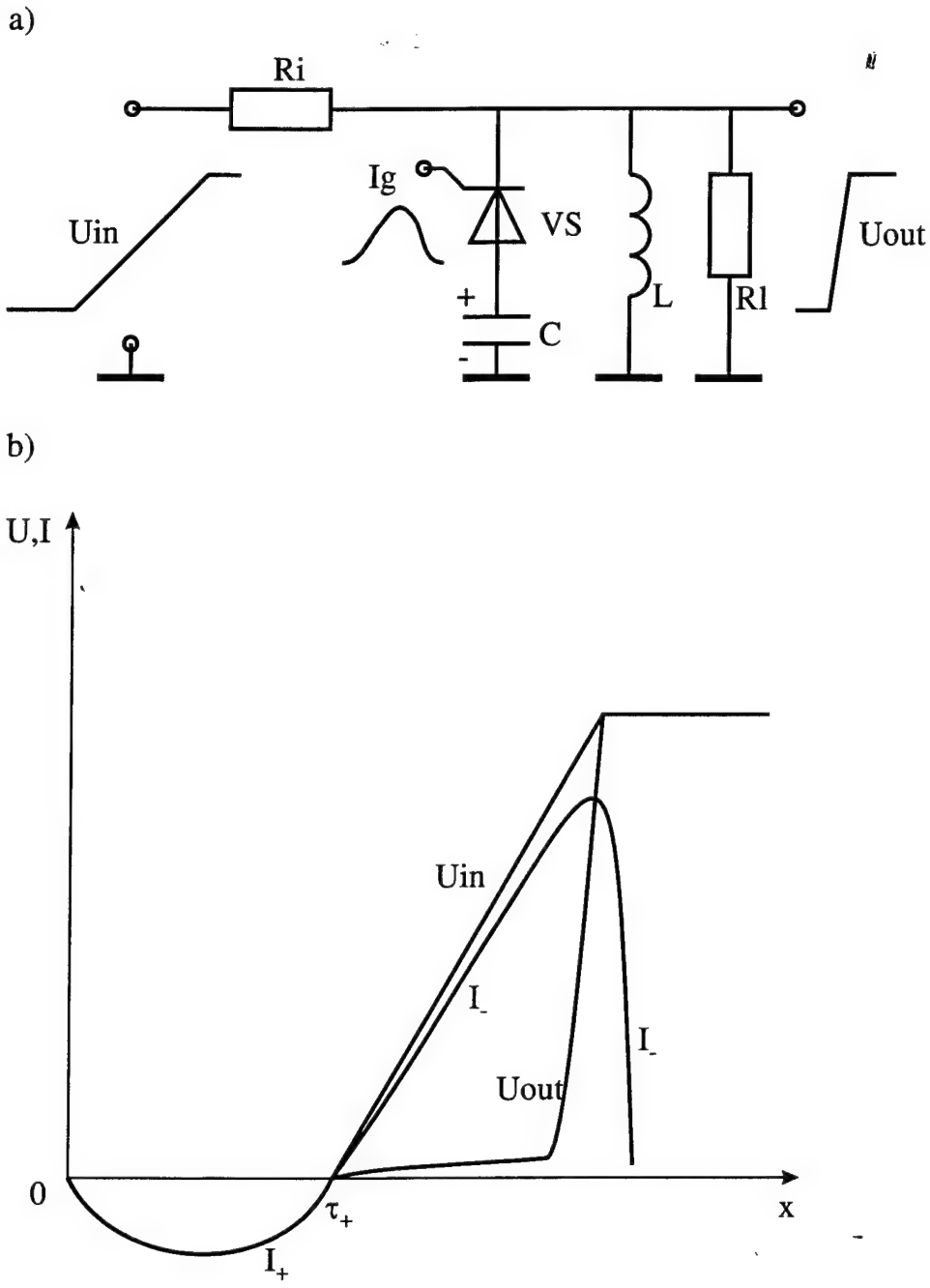


Fig.4.14
Fast restoration in $n^{++}p^{+}np^{+}$ structure.
(circuit, transient voltages, current).

Nearby n^{++} and p^+ emitters two diffusion regions, enriched by nonequilibrium carriers, appear (fig. 4.13). The widths of the regions may be estimated by the same as for diodes equation $L \approx \sqrt{D\tau_+}$ where τ_+ is the halfperiod of the current oscillation of LC circuits. In other parts of p and n layers (not occupied by diffusion regions) the electron transfer is due to electric field force. The carriers distribution closely resembles the distribution in DSRD. The forward current pulse width must be chosen from the well known for DSRD condition $L << w_n$, $\tau_+ \approx w^2/D$.

When the forward (pumping) current runs out, the external slowly rising voltage (U_{in}) of reverse polarity is applied (fig. 4.14). For some period of time (τ_-) the thyristor remains in the high conducting state. The current (I_-) limited by external circuits $I_- = U_{in}/R_i$ extracts the injected carriers.

The processes of plasma dispersion in either diffusion regions (by n^{++}-p and n-p^+ junctions) closely resembles the processes in DSRD as well. Two regions of the reverse gradient of concentration appear. Let us consider the process in n-layer. The main difference from the case of a diode is in the possibility of the existence of only one diffusion region in the layer. Nevertheless during plasma extraction the holes current flows from the p^+-region into the n-region. The diffusion region by n^{++}-p^+ junction plays the role of the second diffusion region by n^+-n junction in the case of p^+-n-n^+ diode (see section 4.2.2).

The holes extracted from the diffusion region in p^+-layer move through p^+-layer and get into the n-layer. The ratio of the charges accumulated during the "pumping" period in these two diffusion regions is the same as in the case of DSRD (see section 4.2.3 p.47). The ratio of extracted charges follows the "rules" of DSRD as well. That means: the made earlier for DSRD (section 4.2.3) conclusion, that the moment of the extraction of all nonequilibrium carriers from the drift regions coincides with the moment of the disappearance of diffusion region is valid for thyristor. That makes possible sharp transition between slow voltage rise at the stage of extraction of nonequilibrium carriers and fast one at the stage of extraction of equilibrium carriers, when well known for DSRD condition is satisfied ($L << w_{SCR}$.)

It should be noted, that in a case of narrow p^+-layer ($w_p << L = \sqrt{D\tau_+}$), the diffusion region in p^+-layer overlaps p^+-layer and occupies a part of n-layer. In this case the difference between the processes in thyristor and diode disappears.

The effect of the super fast restoration has been checked experimentally. The thyristor with blocking voltage > 1.6 kV was connected in the circuit shown at fig. 4.14. The turn off time was less than 4 ns for pumping time ≈ 300 ns and braked current > 50A. In thyristors as well as in transistors the anode current concentrates in the region near gating electrodes. Due to this concentration average anode current must be less than optimal for the fastest switching off and the turn off time is 1.5-3 times more than in DSRD having the same p^+ -n junction.

5. Investigation of properties and limitations of picosecond closing switches based on reversible breakdown of p-n junctions

5.1 General consideration

As was shown in the first chapter, impact ionization is an extremely powerful mechanism for the generation of electron-hole pairs, and can provide enormous growth rates for the concentration and current: $G \approx 10^{32}$ cm^{-3}/s, $j' \approx 10^{20}$ A/(cm^2·s); the direct realization of such a process is possible by means of the application, for example, to a n^+-- n -- n^+ structure of a very short high-power voltage pulse. When the n layer has a thickness of approximately 10^{-2} cm, the amplitude which guarantees the threshold ionization coefficient $\alpha_\infty \approx 10^5 cm^{-1}$, is

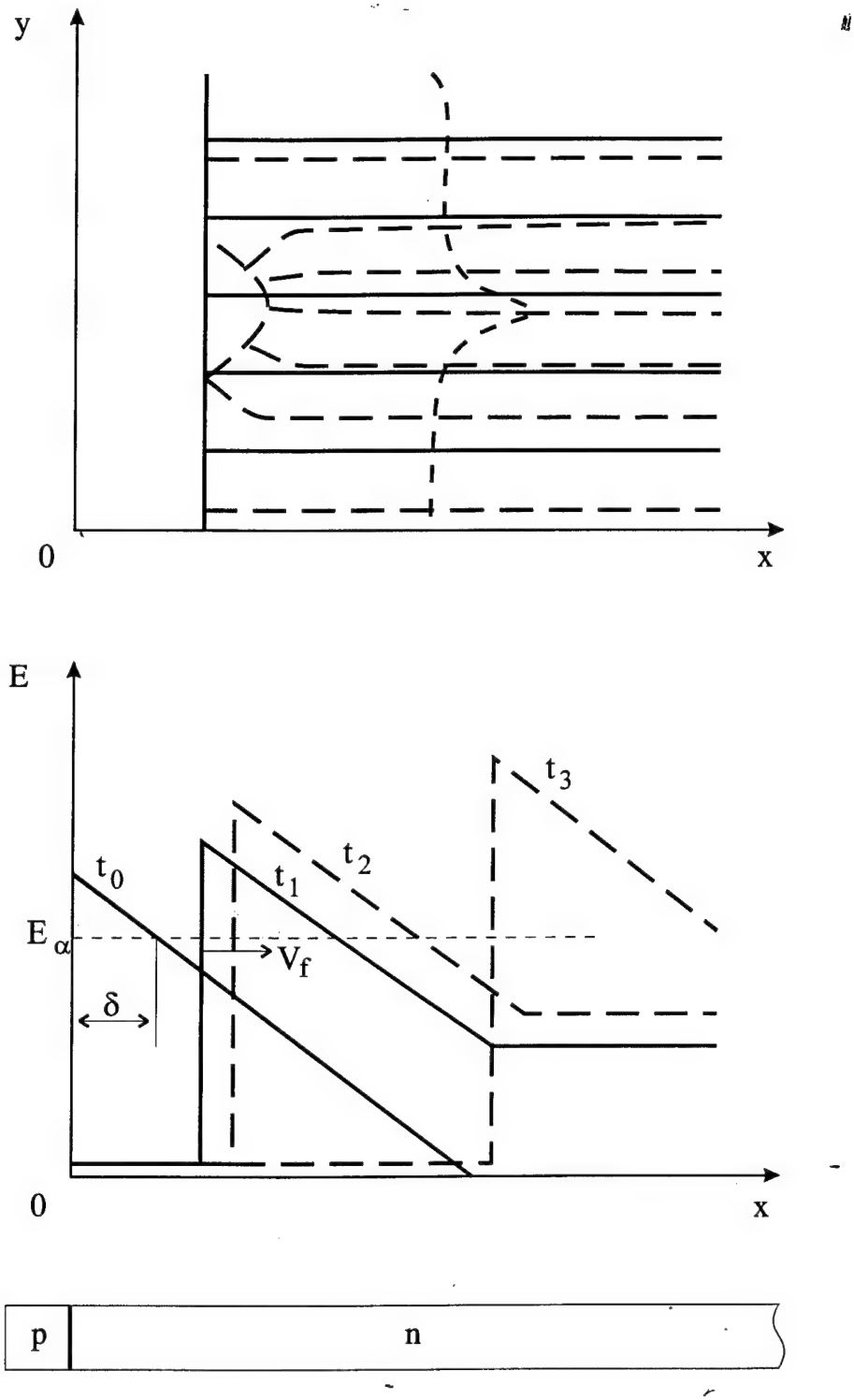


Fig.5.1
Shock ionization wave
The instability of the flat front on an ionization wave.
The dotted lines indicate the perturbed state.

equal to approximately 10^4 V. The characteristic time for the multiplication of the carriers in the case of such a coefficient equals approximately 10^{-12} s. The duration of the pulse should not strongly exceed this value. In the opposite case, due to the increase of the carrier concentration, the conductivity current begins to increase, and, consequently, a larger power is required from the generator. The shaping of such a pulse presents a still unresolved problem, and makes the approach unpromising. Obviously, for practical use it is necessary that the independent fast development of the ionization process be guaranteed when there is a relatively slow or (and) weak triggering source.

Such process of fast reversible breakdown in high voltage p-n junctions [1] was discovered in Russia by the same group of PTI, that discovered fast restoration in high voltage p-n junctions as well. The physical picture of the process has been suggested by the same group and shortly was considered in part 1 (section 1.1, fig. 1.4). This effect has been exploited for subnanosecond switching in Silicon Avalanche Shapers (SAS). Since then SASs have been used for pulse generation by many groups in Russia and in USA. The papers devoted to physical processes of switching of SAS are few in number].

In the next chapters we will consider the physics of the process in more detailes.The consideration is based for the most part on the results of investigations that have been conducted by FTI-MP group. As was shown (Fig1.4) the switching properties of SAS are based on fast ionization shock wave generation and propagation. Fast ionization shock waves have been known in gases for more than half century [13], [14]. The possibility of a wave breakdown in semiconductors and the concept of an ionization wave, was developed for the TRAPATT system of operation for silicon avalanche-flight diodes which are intended for the generation of microwave signals [15] . Let us briefly examine this mechanism.

A diode with a p^+-- n -- n^+ structure is subjected to a constant voltage having blocking polarity (fig. 5.1). The ratio of the thickness of the n layer w and the concentration of the doping impurities in the n layer N_d is chosen so that the n layer is completely covered by the space charge region in the case of when the maximum value for the field intensity near the p -- n transition is lower than the threshold for the impact ionization E_a. The leakage current with density j_0, determined by the heat generation processes, goes through the diode. At a jump during time δt, which is so small that during it the carriers fail to shift noticeably ($\delta t << \tau_s = w/v_s$), we will increase the current density through the diode to value j. Since the conductivity current is small ($j_0 << j$), the field intensity in the n region begins to increase according to the law $E = jt/\varepsilon$, and it will be raised higher than the impact ionization threshold in layer δ. At first, when there was a significantly small j_0 value, the increase in the conductivity current due to ionization will not prevent the reinforcement of the field. Then, after time interval $\Delta t \geq \tau_i \ln(j/j_0)$, where $\tau_i = 1/(\alpha v_s)$, the concentration of the carriers increases so that the density of the conductivity current exceeds the j value, and the field intensity in the δ layer begins to decrease to a value which is less than E_α , or the ionization threshold. It is obvious that the field maximum will shift (fig. 5.1). In the area of the new maximum, due to ionization, the field intensity again decreases, the maximum shifts, and so forth; that is, an ionization wave arises which runs with a speed $v_b > v_s$. On the basis of what has been stated, we get the following evaluation for the speed of the wave:

$$v_f \approx \frac{\delta}{\Delta t} = \frac{\delta \alpha v_s}{\ln(j/j_0)}.$$

On the other hand, taking into consideration that in front of the wave front the current is the displacement current, from fig. 5.1 it is easy to get the velocity of the shift in point d, at which the field intensity has some kind of fixed value, for example E_α:

$$\frac{dE_\alpha}{dt} = \frac{\partial E}{\partial \chi}\frac{\partial \chi}{\partial t} + \frac{\partial E}{\partial t} \equiv 0,$$

where $\partial\chi/\partial t = v_b$; $\partial E/\partial\chi = -qN_d/\varepsilon$; $\partial E/\partial t = j/\varepsilon$.

From here it follows that

$$v_f = j/(qN_d) = v_s j/j_s.$$

The equations above coincide with those obtained by the authors of references [13, 14, 15].

Thus, in the case of a current jump after a short period of increase in the voltage Δt, the movement of the ionized wave begins, and the voltage on the diode drops during time $\tau_\varphi = w/v_b$ almost to zero. It should be noted that the ionization wave of a TRAPATT system is analogous to "the wave of the potential gradient" [13, 14] in a gas, and the expressions for the speed of these waves are identical with respect to form . The difference lies in the physical principle which creates the potential gradient. In the case of a TRAPATT system, this is the space charge of the immobile ionized donors, and in the second case, these are the geometrical distortion of the field or the fields of a little-mobile "ion island" [13,14].

The examined mechanism for the generation of plasma has found wide enough use in avalanche flight diodes which are used for the generation of microwave oscillations. However, for purposes of the fast switching of large powers, such a mechanism has little effectiveness. The reason is as follows. In the case of the fast switching of a power P_k, that is, during switching of a device in the conducting state during small amount of time τ_φ , it is desirable that the trigger pulse from the external additional source have either a small power ($P_y << P_k$) or a large increase time ($\tau_y >> \tau_\varphi$). In other words, the switch should, as a minimum, guarantee a gain either with respect to power or with respect to fast response. From what has been stated previously, it follows that in TRAPATT diodes, the gain with respect to quick response $\kappa = \Delta t/\tau_\varphi \approx \delta/w \approx 1$. The gain with respect to power is equal to the ratio of the constant shift voltage to the amplitude of the super-voltage pulse, and it is also small (approximately 1) . The lack of a gain with respect to fast response and power dies not prevent the use of the TRAPATT diode in microwave generators for which a fast voltage drop on the diode, formed during the wave run, "leaves" for the external circuit and almost completely "returns" with a change in the phase to the diode for starting up the wave to the next cycle. But the lack of a gain makes the direct use of the process described above unpromising for purposes of fast commutation.

The main difference between TRAPATT waves and waves in SAS is in the initial conditions. In TRAPATT case there is constant flow of primary carriers in the time of field rise. In SAS case the field rises when there is no primary carriers to start ionization in the high field region although the field intensity is high enough for very quick breakdown. The ionization is delayed until carriers arrive from low field region. It allows to call the ionization in SAS "delayed ionization" and the high field region is "overvoltaged".

Different modes of ionization shock waves, mentioned above, have some common in features. We will begin from considerations of such "common" for all modes processes and then consider the processes that are special feature of SAS operation.

5.1.1 Ionization wave front stability.

The problem of the stability of a flat ionization wave front should be discussed here. As is obvious from the mechanism for the movement of a wave, its speed is determined by the ionization rate of the front: the higher the ionization rate, the greater the speed. Let us examine the movement of a flat front (line of the maximum of the voltage intensity), that is, a one-dimensional pattern (fig. 5.1); initially the field is uniformly spread across the area of the diode. Let at the wave front arise a disturbance in the form of a "protuberance". It is obvious

that such a distortion of the front line causes the thickening of the field's lines near the "protuberance", that is, it causes an increase in the field intensity. This effect of reinforcing the field at the point of the electrode is well known in electrotechnology. A local increase in the field intensity at the apex of the "protuberance" leads to an increase in the ionization rate and the velocity of motion of the apex in comparison with the remaining portion of the protuberance. That is, the "height" and curvature of the latter increase. This feature, in its turn, again leads to the strengthening of the field at the apex, to an increase in the velocity of motion of the apex, etc. Thus, a wave front which is flat at the beginning should fall apart into "threads" moving with a great speed, which then also determine the increase in the current.

The examined mechanism for the formation of the instability is in essence equivalent to the mechanism for the development of instability in a gas breakdown.

5.1.2 Plasma concentration.

Let us make for ionization waves some evaluations which are general in character, and are not connected with a specific method for exciting a wave.

When filling the single volume of a semiconductor with electron-hole plasma, due to the impact ionization process, the energy going for ionization is drawn up from the initial energy of the electrical field in this volume. As is well known, the density of the field energy $W = \varepsilon E^2/2$, and the change in the energy density when changing the field intensity at dE equals

$$dW = \varepsilon E dE. \tag{5.1}$$

Taking the dissipation of the energy due to all possible collisions into consideration, for one act of ionization, average energy equaling $W_u = qE/\alpha$ is expended (where α is ionization coefficient). Therefore, in order to change the number of pairs which come to be due to ionization with an expenditure of energy dW, we get the following based on expression (5.1)

$$dn = \frac{dW}{W_u} = \varepsilon\alpha dE/q. \tag{5.2}$$

When integrating equation (5,2) within the limits from the initial value at the wave front for the field intensity (E_m) to the final at the "tail" of the wave value (accepted to be equal to zero), we find for $\alpha = \alpha_\infty \exp(-b/E)$

$$n_m = \frac{\varepsilon\alpha(E_m)E_m}{q}\left(\frac{E_m}{b}\right)., \tag{5.3}$$

where n_m is the plasma concentration at the wave tail after filling the volume with plasma.

Let us emphasize that expression (5.3) has a universal character which does not depend on the velocity of motion of the wave front.

The field intensity drop behind a passing ionization front with a width Δ is connected with the fact that the electron-hole pairs which come to be due to ionization are separated by the field, and a space charge arises. Taking into consideration that the fraction of the separated carriers of the complete number at the front is approximately v_s/v_φ, we get the following from the Poisson equation

$$E_m \approx \frac{qn_m\Delta v_s}{\varepsilon v_n}. \tag{5.4}$$

When substituting expression (5.4) here, we find

$$\Delta \approx \frac{1}{\alpha_m}\left(\frac{b}{E_m}\right)\frac{v_n}{v_s} \tag{5.5}$$

which is the evaluation for the dimensions of the ionization wave front.

5.1.3 The breakdown delay when there are no primary carriers

In order to generate fast ionization wave in the case of delayed ionization, there should first of all be determined under what conditions the creation in the semiconductor of a overvoltaged region is possible, and whether it is possible at all the existence of the region

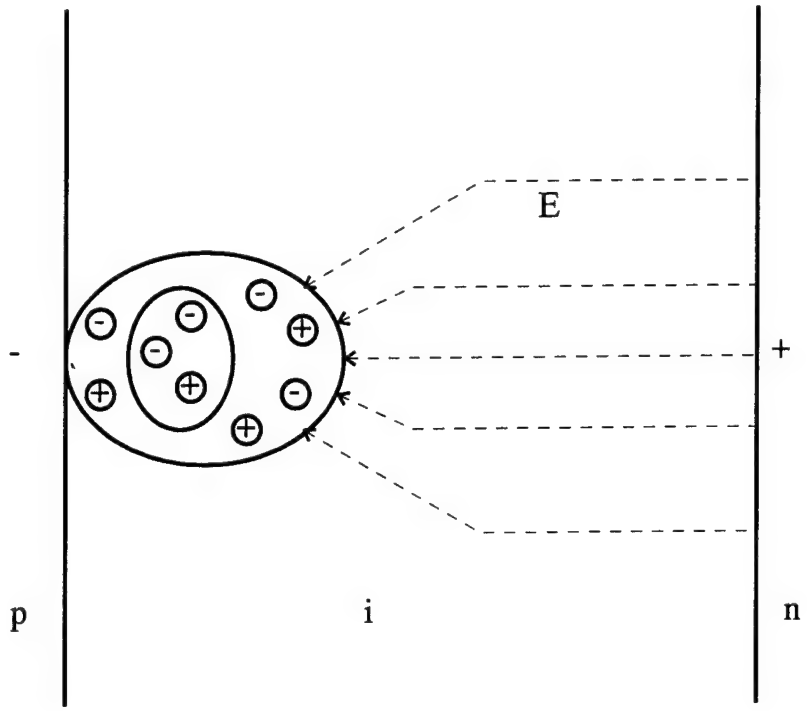


Fig.5.2
The evaluation of the current of a single avalanche.

w_α in which the ionization integral exceeds one: $\int_{w_\alpha} \alpha_i d\chi >> 1$, where $\alpha_i = \alpha_n \exp[-\int_{w_\alpha}(\alpha_n - \alpha_p)d\chi]$ is the effective ionization coefficient; α_n and α_p are the ionization coefficients for the electrons and the holes. Let us make note of the fact that $\int \alpha_i d\chi = 1$ is the condition for stationary breakdown . Obviously, the realization of the overvoltaged condition is absolutely necessary for any ionization process leading to the filling of the volume of the semiconductor with plasma during an amount of time which is less than the flight time of the carriers through it.

Let a leakage current having density j_0 flow through a p^+-- n transition. Then the SCR boundary for time $T = q/(j_0 S)$ intersects one carrier at an average. The flight time of a carrier through the SCR is $\tau_s = w_{SCR}/v_s$. From here for $T > \tau_s$, we find the average time during the course of which the volume of the SCR will be completely free of charge carriers:

$$T_{fr} = T - \tau_s = \frac{q}{j_0 S}(1 - \frac{j_0 V}{q v_s}), \qquad (5.6)$$

$\Pi_{\xi\eta}$ where V = SwSCR is the volume of the SCR.

It is easy to evaluate that in structures with a leakage current through a p^+-- n junction $j_0 < 10^{-9}$ A/cm^2, the value T_{fr} lies in the nanosecond range when $S < 0.1$ cm^2, $w_{SCR} = 0.01$cm. Thus, when supplying a voltage pulse with a front duration of $\tau_\varphi < T_{cp}$ to the structure, it is possible to realize a situation when the field intensity significantly exceeds the threshold for impact ionization, but on the average it can be considered that there will be no acts of ionization. However, the moment that an ionization initiating carrier gets into the SCR is not synchronized with the moment that the voltage is increased. Therefore, there exists the probability that the carrier gets into the SCR all the same at that moment when the field intensity in it becomes greater than the impact ionization threshold. Obviously, as a result of the chain of ionization acts initiated by these carriers, there arises a single conducting channel, or avalanche, which will expand along the field lines at a rate v_a (see fig.5.2) . The current caused by the appearance of such a channel can be evaluated with a large reserve from above for the following suppositions: the conductivity of the channel is infinite as it moves through, the channel expands spherically (fig. 5.2), and the voltage on the structure (U_0) is constant. In this case, the evaluate comes down to a simple problem on the displacement current which arises during the expansion of the conducting sphere with a radius $R = v_a t$. Its solution has the following appearance:

$$I_0 < \frac{\varepsilon v_a U_0 w}{(w - v_a t)^2}, \qquad (5.7)$$

where t is the time read out from the time the carrier lands in the SCR; w is the thickness of the n layer.

In silicon, the coefficient for radiation recombination is small, and it should be expected that, at least for the initial stage of the development of the avalanche, while its volume is small, its radiation is weak. Therefore, it is possible to disregard photogeneration in front of the head portion of the avalanche.

As is well known, the frequency of tunnel ionization for an electron can be presented in the following form : $w_n \approx w_\tau \exp(-E/E_\tau)$, where for silicon $w_\tau \approx 10^{13}$ s^{-1}; $E_\tau \approx 4 \cdot 10^7$ V/cm. For an average value $E \approx 4 \times 10^5$ V/cm, corresponding to the two times the overvoltage on the diode, we get $\exp(-E/E_\tau) \approx 10^{-44}$; that is, it is possible to disregard tunnel ionization. The reinforcement of the field in the head portion of the avalanche can increase the role of the tunnel ionization, but still the exponent remains large. However, the extremely strong dependence of the ionization rate on the field intensity does not make it possible to make a more accurate evaluation of the effect of tunnel ionization.

The problem of the velocity of motion of the avalanche front when there is no photo- and tunnel ionization, and also when there is no meeting flow of ionizing carriers, requires

some discussion. The electron which comes to be at the wave front is accelerated into a strong electrical field up to the moment of collision at the optical phonon, like a free one. After collision it completely loses the momentum. As a result of the collision , an average speed equal to the saturated speed v_s is established. It is a small number of carriers which avoided acts of collision , and who pass through the entire path $\lambda_i = 1/\alpha$ up to the accumulation of ionization energy E_i , like free ones, leads the impact ionization . In agreement with these, the average speed along the entire path for the fast electrons is

$$v_{cp} = \frac{1}{2}v_m = \frac{1}{2}\sqrt{\frac{2E_i}{m}} \qquad (5.8)$$

and can be higher than the saturated speed; for example, for silicon, taking into consideration the law for momentum conservation, $E_i = 1.8$ eV , which, in agreement with expression (5.8),(5.8) gives $v_{cp} = 4.7 \cdot 10^7$ cm/s; that is, the four-fold surpassing of v_s .

However, the fraction of such fast electrons from their total number equals $n_0/n \approx \exp(-b/E)$, where $b \approx E_i/(q\lambda) \approx 1.2 \cdot 10^6$ V/cm (λ is the length of the electron's free run before collision with an optical phonon).

Let us recall that $\alpha_n = \alpha_\infty \exp^{(-b/E)}$, $\alpha_\infty = 0.65 \cdot 10^6 cm^{-1}$ [16]. Thus, when E << b, the fraction of fast electrons is small ($5 \cdot 10^{-2}$ when E≈ $4 \cdot 10^5$ V/cm), and they only somewhat widen the front in comparison with evaluation (5.5). The main mass of electrons (>95%) moves with the saturated speed Taking what has been stated into consideration, it is possible to assign $v_a = v_s$. Then the evaluation according to formula (5.7) for real values $w \approx 3 \cdot 10^{-2}$ cm, $U_0 = 10^3$ V yields $I_0 < 10^{-2}$ A when $t < 10^{-9}$ s. Such a small current when there is typical resistance of the external circuit of the diode of 50 ohm cannot remove the voltage from it and cannot prevent the formation of a overvoltaged region. Then the probability that several carriers will get simultaneously exponentially decreases with an increase in their number. The maximum area of the "trail" from one such electron is less than $w^2 = 10^{-4}$ cm^2, therefore, even when the area of the entire structure is 10^{-2} cm^2, 99% of the entire area of the "overvoltaged" region will be free of carriers.

Thus, in a structure with a small leakage current, it is possible to expect a delay in ionization for a time which is close to the flight time.

The possibility of such a delay was verified experimentally on a pin diode created by the diffusion of aluminum and boron in n- silicon with a specific resistance $\rho = 270$ ohm·cm. The thickness of the i layer $w_i = 120$ μm, and the depth of the diffusion of aluminum equaled 100 μm when its surface concentration was 10^{17} cm^{-3}. The n^+ type layer was formed by the diffusion of phosphorus through the polished surface, which ruled out the possible injection of carriers from a contact. The diode had a stationary breakdown voltage of 2800 V. The measurement of the volt-farad characteristics showed that the complete overlapping of the i layer by a space charge region occurred at a voltage of 200 V, equal to that calculated for the given w_i and ρ values. The short (3ns) high voltage pulse was applied to the diode in addition to constant bias $U_0 = 1000$ V. The measured current trace showed that, in the case of a complete voltage on the diode of approximately 5000 V, only a displacement current went through the diode, (the curve of this current almost repeated the curve of the rate of change of the voltage); that is, the conductivity current was negligibly small (less than 1 A). Such a "currentless" state exists for 3 ns. When there is a further increase in the voltage by 10%, an irreversible breakdown occurs. Let us evaluate the degree of overvoltage intensity.

Assuming the exponential distribution of the concentration of aluminum in the p layer ($N_{Al} = N_s \exp^{(-x/a)}$, it is easy to determined, while integrating the Poisson equation twice, the voltage drop on the section of the SCR which is positioned in the p layer:

$$U_p = a(E_n - qw_pN_d/\varepsilon). \qquad (5.9)$$

The characteristic length of the doping profile is determined from the condition $a \approx w_p/[\ln(N_S/N_D)] = 1.16 \cdot 10^{-3}$ cm, where N_d is the concentration of the impurities in the i layer. It is easy to show that $U_p = 250$ V, and that the voltage drop on the i layer $U_i = 2550$ V for the threshold of the stationary breakdown corresponds well with the results of the stationary breakdown threshold.. The maximum field intensity in this case equals $2.3 \cdot 10^5$ V/cm on the p - i junction, and it decreases to $1.95 \cdot 10^5$ V/cm on the i - n junction. The corresponding time for electron ionization $\tau_n = (\alpha_n v_s)^{-1}$ equals approximately 40 and 100 ps. When there is a total voltage (pulse plus constant bias) of 5 kV, the following will hold: $U_p = 400$ V, $E_m = 4 \cdot 10^5$ V/cm, $\tau_n = 3$ ps at the p - i junction, $E_m = 3.6 \cdot 10^5$ V/cm, $\tau_n = 4$ ps at the i -- n junction; that is, in the overvoltaged state, the ionization time decreases by more than an order of magnitude as compared with the corresponding stationary breakdown. If it is assumed that the distribution of the carriers which create the initial leakage current with density j_0 is homogeneous with respect to volume, then when there is such a small ionization time, for flight time $\tau_s \approx$ 1.5 ns (taking the expansion of the SCR into consideration), the leakage current should have increased by $\exp(\tau_n/\tau_s) \approx \exp^{350}$ times! In other words, when there is any small density for the leakage current j_0, after time τ_s the complete current should have surpassed the 1 A level estimation for maximum possible conductivity current noted above.

It should be noted that, when there is such an increase, then the avalanche created by a single electron also should make the transition to the streamer stage long before approaching the anode. The limitation of the current in this case is evidently connected with the "damping" effect of a thick p layer with a relatively low concentration $N_.$. When there is a large current density in the streamer, the voltage drop on the p layer increases, and the field in the streamer is weakened. Since the thickness of the p and i layers are almost identically, such a redistribution can lead to the two-fold decreasing of the average field intensity, even to the stationary breakdown level. After accepting a current density in a streamer of 10^6 A/cm^2 and its cross section 10^{-4} cm according to the data in reference [18], we find that the complete current of a single streamer equals approximately 10^{-2} A, and this is too small for the redistribution of the voltage between the diode and the resistance of the external circuit (100 ohm).

On long pulses (50 ns), there was a failure to create a significant overvoltage due to the irreversible breakdown of the diodes.

Thus, in semiconductor silicon diodes, for a short time (approximately the flight time), the creation of overvoltaged regions, in which the characteristic ionization time is more than an order of magnitude lower than the ionization time corresponding to the threshold of the stationary breakdown, is possible.

5.2 The effect of super-fast switching in diode silicon structures

5.2.1 The main features

As was shown above, in pin structures, it is possible, at least for a time equal to the flight time, to create a region of a strong electrical field in which each carrier can give rise to many (more than dozens) of secondary carriers during the flight time. In this case, the total voltage on the diode significantly surpasses (by approximately 2 times) the stationary breakdown voltage, and the conductivity current is extremely small. When there is an insignificant increase in the duration of the pulse, the conductivity current sharply increases, which is accompanied by the complete loss by the diode of its blocking capability; that is, an irreversible breakdown occurs.

However, similar experiments on a super-voltage on p^+-- n -- n^+ structures have shown the possibility of a reversible breakdown: after a certain time after the breakdown, the

blocking capability of the structure is completely restored. In fig. 1.3 was shown that the voltage on the structure, after reaching the maximum value $U_m = 3$ kV after delay time $\tau_3 = 2$ ns, drops extremely quickly (in τ_φ <0.2 ns) to the residual value $U_{min} = 200$ V, and then slowly (in approximately 10 ns) increases to the stationary value $U_* = 2.1$ kV.

The cooling of the structure from room temperature to the temperature of liquid nitrogen somewhat decreased the delay and the switching voltage.

Let us make note of the main properties of the features of super-fast switching.

1. Extremely high stability of the switching moment relative to the moment for the beginning of the increase in the voltage (jitter within the limits of 30 ps is the value determined by the instability of the oscilloscope triggering).

2. Large ratio of dU_a/dt applied to the diode and dU_b/dt during switching on, called-gain ($k = \frac{dU_a}{dt} / \frac{dUb}{dt} \approx 10$).

3. Current density after switching significantly exceeds the density of the threshold current, which can be guaranteed by equilibrium carriers in the n base; that is, the entire n layer is strongly enriched by carriers in the switching process.

4. The switching time $\tau_\varphi < 0.2$ ns, which is an order of magnitude less than the drift time of the carriers through a modulated volume with the saturated speed.

5. The switching voltage $U_m = 3$ kV; that is, it is twice as large as the stationary breakdown voltage $U_n = 1.5$ kV.

As was already noted, the wave ionization processes during a breakdown are continuously connected with the mechanism which guarantees the existence of primary carriers which initiate ionization. A number of factors bear witness to the fact that the processes for the thermal generation of carriers, both in the SCR and in neutral regions, cause a leakage current of p -- n transitions to the breakdown, do not play a part in the case which interests us. The following for example, are belong to this fact:

1. Very large gain ($\kappa \approx 10$).

2. Weak dependence of the switching process on temperature: at the very same time the leakage currents caused by thermal generation decrease by many orders of magnitude (approximately 10^{30}!) when there is a change in temperature from room temperature to the temperature of liquid nitrogen (77 K), the delay in ionization does not only not increase, but also decreases somewhat.

3. The extremely high stability of the delay, which is retained even at the temperature of liquid nitrogen. As may easily be shown taking into an account the famous Shockley -- Sa -- Nnice theory, at room temperature the density of the leakage current, caused by thermal ionization, equals 10^{-9} A/cm^2 for the studied structures. It should be noted that in the experiments, due to the small area of the structure, there was a failure to get rid of the surface leakage currents, and it was only possible to evaluate the upper boundary of the leakage current. At the temperature of liquid nitrogen, the minimum recorded currents were less than 10^{-11} A when there was a displacement of 1 kV. Even at these currents, the average time between moments of the passage of carriers through the plane of the p -- n junction should equal $\delta t \approx q/I \geq$ 10ns. In agreement with this, the stability should be close to the value δt.

Since the switching process does not depend on frequency in a very wide range of pulse tracking frequencies (from tens of kilohertz to single pulses), the effects connected with the accumulation of residual carriers in strongly alloyed layers from cycle to cycle (the storage effect in the TRAPATT mode]) can also be ruled out. Consequently, it is necessary to assume a purely field mechanism or field mechanisms for the generation of primary carriers which initiate impact ionization.

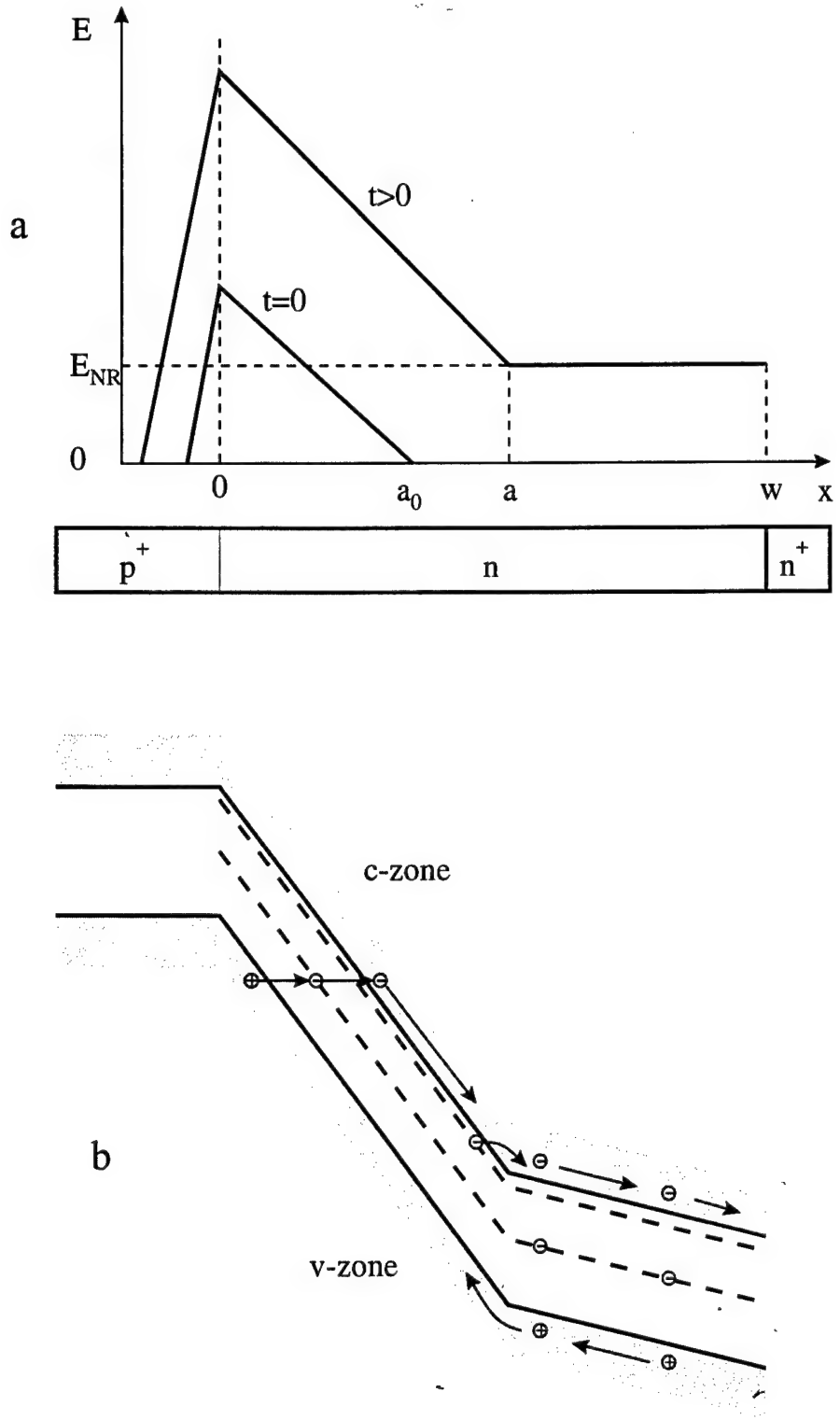


Fig.5.3
The distribution of the field (a) and potential (b)
in a p^+nn^+ diode when there is a rapid increase in the voltage.

Obviously, when there is the complete switching time $\tau_\varphi = 0.2$ ns, the average distance between the primary carriers in the super-voltage region should be no more than $l \approx \tau_\varphi v_s \approx 2 \cdot 10^{-3} cm$. The concentration $n_0 = l^{-3} \approx 10^8$ cm^{-3} and leakage current density $j_0 \approx qn_0v_s = 10^{-4}$A/cm^2 correspond to this l value. (Let us recall that the nonlocal effects mentioned in section 5.1 will have an effect when there are fields with an intensity of approximately 10^6 V/cm).

The following experiment yields an evaluation of the leakage current (upper boundary), at which the formation of a overvoltaged region is still possible). In the tested diodes from the side of the n^+ layer in the contact, a "window" with a diameter of 1 mm, and with a depth of 30 μm in the silicon, was etched. This area was illuminated by a focused beam from a lamp. The constant bias voltage equaled 500 V. The currents for various values of the intensity of the illumination had been measured. Then fast rising voltage was applied to the diode for the each illumination measured before. It was shown that a photo- current of 10^{-8} A noticeably decreased the delay in the current growth and the switching voltage, and increased the switching time to 0.5 ns. When the current was 10^{-5} A, the switching effect disappeared completely.

We have change p-n junction leakage current by heating as well. The results are the same as when the diode is illuminated - the origin of the initial (leakage) current plays no role. It is the value of the leakage current what is important..

In agreement with what was stated above, the observed switching process is conveniently divided into the following stages: the creation of a overvoltaged region in the absence of initiating carriers; the generation of initiating carriers; the excitation of a fast ionization wave; the extraction of the plasma accumulated after the wave passes. Let us examine these processes in more detail.

5.2.2 The dynamics of the overvoltaged region

Let us examine the distribution of the field in a p^+-- n -- n^+ structure, which contains an n region with thickness w and is connected in series with resistance R_l to the generator of voltage U_r , increasing linearly from the initial stationary value U_0 with rise rate U'; that is $U_r = U_0 + U't$. At the initial stage, the distribution of the field in the SCR with dimensions $a_0 = (2\varepsilon U_0/qN_d)^2$ is linear. In the neutral region (NR), the field intensity equals zero. The increase in the voltage on the structure leads to an intensification of the field in the SCR and its expansion. The latter is connected with the drift of the main carriers in the NR under the effect of the field, which is caused by the flow of the current through the structure; the current is a displacement current in the SCR, but in the NR it is partially a conductivity current. Since the initial field intensity E_{NRO} is equal to zero, then during the course of a certain initial time interval, it remains small and it fulfills the condition for the constancy of mobility, and, consequently, $v_n = \mu_n E_{NR}$ as well. Then (when there is a sufficiently large rate increase in the voltage), the field intensity can reach a value of $E_{NRO} > E_s$, at which the drift speed becomes saturated. This problem on the distribution of the field at the given voltage is close to the problem previously examined in point 4.2.3.(fig.4.7 and fig.5.3).

When using the very same approach as that when deriving formula (4.34), we get for condition $v_n = \mu_n E_{NRO}$ the following equation in dimensionless variables which are more convenient in the given case:

$$\beta\varphi_{\tilde{t}\tilde{t}} + \varphi_{\tilde{t}} + \gamma\varphi^2 + \varphi = \tilde{t}\,, \qquad (5.10)$$

where

$$\beta = \frac{\tau_0 a_0^2}{\tau_M w^2 \psi^2},\ \tau_a = \frac{R_H S_E}{a_0};\ \gamma = \frac{wU'\tau_M\psi}{4a_0U_0}.$$

The expressions for τ_c, τ_M, ψ, and φ are the very same as in formula (4.34). The following values are accepted as units: $t_u = \tau_M w\psi/a_0$; $E_u = U'\tau_M/a_0$. The initial values are zero.

Let us evaluate the coefficients in equation (5.10). For the case which interests us ($U_0 = 10^3$ V; $U' = 10^{12}$ V/s; $N_d = 10^{14}$ cm^{-3}; $w = 2.5 \cdot 10^{-2}$ cm; $a_0 = 10^{-2}$ cm; $S = 0.05$ cm^2; Rl= 10^2 ohm), we get $\tau_M = \rho\varepsilon = 5 \cdot 10^{-11}$ s; $\tau_s = \varepsilon S R_H/w = 2 \cdot 10^{-10}$ s; $\tau_a = 5 \cdot 10^{-10}$ s; $\psi = 5$; $\beta = 6 \cdot 10^{-2}$; $\gamma = 0.15$; $t_u = 0.6$ ns; $E_u = 5 \cdot 10^3$ V/cm. Taking the smallness of the γ parameter into consideration when there are rather small φ values (such that $\gamma\varphi < 1$), in equation it is possible to disregard the squared term. Physically, such an approximation denotes the case of the smallness of the displacement of the SCR boundary from its initial value a_0. In this case, the equation is easily solved:

$$\varphi = C_1 \exp(r_1 \tilde{t}) + C_2 \exp(r_2 \tilde{t}) + \tilde{t} - 1, \tag{5.11}$$

where

$$C_1 = -\frac{1+r_2}{r_1 - r_2};\ C_2 = \frac{1+r_1}{r_1 - r_2};\ r_{1,2} = \frac{-1 \pm \sqrt{1-4\beta}}{2\beta}$$

Taking the smallness of β into consideration, and disregarding the quickly attenuating exponent with r_2 factor, we simplify the solution to (5.11) and we get in dimensional form:

$$E_{NR} = E_s \frac{U'}{U'_K}\left[1 - \exp(-\frac{t}{t_u})\right], \tag{5.12}$$

where $U'_K = E_0 v_s$ is the critical rate of growth of the voltage, $E_0 = qN_d a_0/\varepsilon$ is the initial field intensity when $\chi = 0$.

From (5.13) it follows that the field intensity in the NR can reach value E_s only in the case of condition $U' > U'_K$. From expression (5.12) it is easy to determine the moment at which the drift speed of the electrons reaches saturated value v_s:

$$\tau_0 = \frac{\tau_M \psi w}{a_0} \ln\left(1 - \frac{U'_K}{U'}\right). \tag{5.13}$$

When integrating expression (5.12), it is possible to find the position of the boundary of the SCR to this moment (fig.5.3):

$$a = a_0 + t_u \frac{U'}{E_0}[t/t_u + \exp(-t/t_u)] \tag{5.14}$$

In this very approximation, taking into consideration that $w\psi/a_0 >> 1$, we get the change in the current density in time when $t < \tau_0$:

$$j = \frac{U'\varepsilon}{a_0}[1 - \exp(-t/t_u)]. \tag{5.15}$$

As follows from the solution of (5.11), the disregarding of the squared term is correct up to moment in time $t < 4U_0/U'$.

For moments in time which are greater than τ_0, using the very same method by which equation (5.10) was obtained, it is possible to construct an equation for E_{NR} when ENR > E_s:

$$\tau_0 \frac{dE_{NR}}{dt} + E_{NR} = E_s + \frac{U_r - U'_K t}{w} - \frac{U'_K v_s t^2}{2a_s w} - \frac{U'_K \tau_c}{a_s w}, \tag{5.16}$$

where U_r is the voltage of the source, the read out of the time is done from moment $t = \tau_0$; $a_s = a(t = \tau_0)$.

For simplicity's sake, let us examine a case with a small time τ_c, that is, with a small load resistance or small area of the structure, when the voltage of the source almost coincides with the voltage on the structure. In this approximation from formula (5.16), we obtain an expression for the linear growth of the voltage U_r.

$$E_{NR} = E_s + \frac{|U' - U'_K - U'_K v_s t/(2a_s)|t}{w}, \tag{5.17}$$

that is, the field intensity in the NR, changing according to parabolic law, increases at first from value E_s, and then, after reaching the maximum value E_{NRm} at moment τ_m, it drops. From expression (5.17), it follows that:

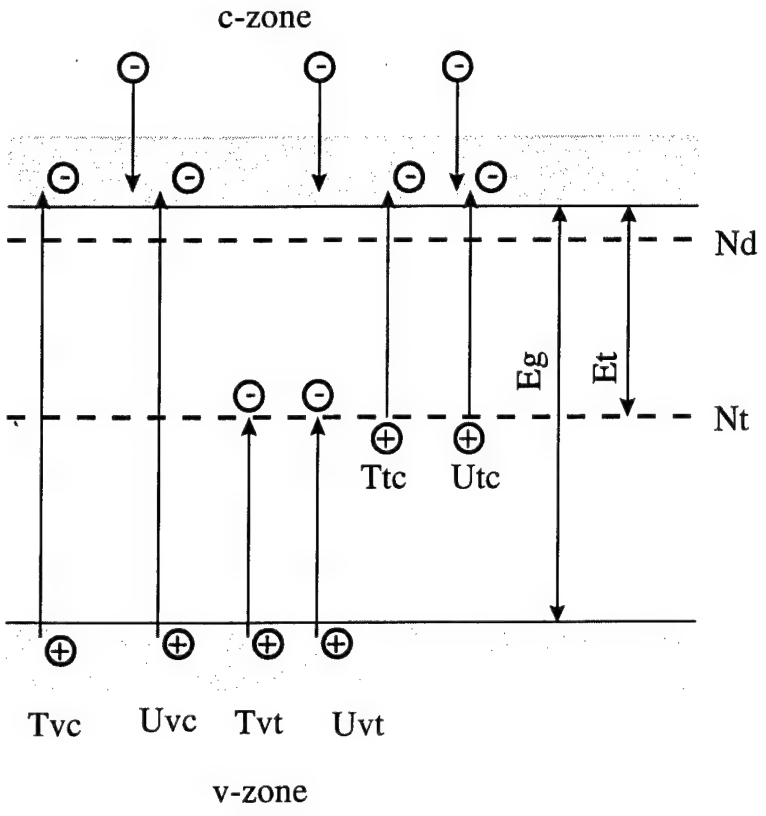


Fig.5.4
Electron transitions during ionization.
Nd- shallow donor level; Nt- deep donor level.

$$E_{NRm} = \frac{U'_K a_s}{2v_s w}\left(\frac{U'}{U'_K} - 1\right)^2, \quad \tau_m = \frac{a_s\left(U' - U'_K\right)}{U'_K v_s}; \tag{5.18}$$

$$j = j_s + \frac{\varepsilon\left(U' - U'_K\right)}{w} - \frac{U'_K v_s t}{a_s w}. \tag{5.19}$$

A comparison of expressions (5.19) and (5.15) shows that, at the first stage ($t < \tau_0$, $ENR < E_s$), the current increases, and at the second stage ($t > \tau_0$, $E_{NR} > E_s$) it decreases. The current maximum should be observed at moment $t = \tau_0$. The section of the current drop in the case of linear or super-linear increase in the voltage bears witness to the saturated drift speed of the NR, since, although the current itself drops, the field intensity in the SCR continues to increase.

It should be noted that the values for the density of the currents, which are determined at moment τ_0 according to expressions (5.15) and (5.19), do not coincide. The relative difference between them equals approximately $\varepsilon(U' - U'_K)/(j_s w)$.

The approximations made during the derivation of expression (5.15) actually come down to the condition

$$j_s >> \varepsilon(U' - U'_K)/w, \tag{5.20}$$

that is, the mèntioned relative difference is small, and it lies below the limit of accuracy of expression (5.15). If $U' >> U'_K$ and condition (5.20) is not fulfilled, then $\tau_0 < \tau_M$ and the stage of the increase in the intensity to value E_s is very short, and the transition to the change in current according to law (5.19) can be considered to be instantaneous.

Let us recall that the conclusion made about ENR = Es at the moment of the current maximum is correct only for the linear law for Ur increasing the current. In an experiment, as a rule on the initial stage of the increase in the current, the rate of growth of the voltage U' smoothly changes from zero to the maximum. If the maximum of the current arrives at this initial section of the change in U' , then the conclusion about reaching $E_{NR} > Es$ is, of course, incorrect.

The dependencies of the current, voltage, and E_{NR} on time, measured experimentally, for moderate values $U' = 2.5 \cdot 10^{12}$ V/s with an error in the 30% limits follows expressions (5.12), (5.15), (5.17), and (5.19).

The maximum field intensity on the p -- n junction can easily be determined in the following manner:

$$E_m = ENR + qN_d a/\varepsilon, \tag{5.21}$$

When $E_m >> E_{HO}$, it is possible to disregard the first component in expression (5.21), and, as follows from formulas (5.14) and (5.21), when $t < \tau_0$, the field intensity increases almost linearly in time. When $t > \tau_0$, the linearity is retained, since $a = a_s + v_s t$, but with a decreased slope.

5.2.3 The generation of flows of initiating carriers

It was shown above that when there is a fast increase in the voltage in the n layer of the diode structure, a strong electrical field ($E > 10^4$ V/cm) can exist. In the NR, it is distributed homogeneously, and in the expanding SCR, it changes linearly with the coordinate. In this case, the current density can be 10^2 A/cm^2 and more. At the present time, a large number of purely field mechanisms for the generation of carriers is known. These mechanisms, generally speaking, are possible both in an SCR, and in an NR under these conditions.

In the general case, the layer of the semiconductor, for example the n layer, contains an doping impurities with "shallow" energy levels which determinesthe initial concentration of the carriers and the type of conductivity, and also a number of deep levels. For illustration, let us examine only one deep donor level (fig. 5.4). In the initial, field-less state, the

concentration of electrons n_0 in the NR conductivity zone is almost equal to the concentration N_d of the shallow donor level. However, at the donor level, a certain number of electrons n_2 can still be found which, as is well known can be evaluated using the following model:

$$n_d = \frac{N_d n_0}{N_c \exp\left(\frac{-qE_d}{kT}\right)}, \qquad (5.22)$$

where N_c is the density of the states; E_d is the ionization energy.

For phosphorus with $E_d = 0.045$ eV at room temperature and $N_d = 10^{14}$ cm^{-3}, the concentration $n_d \approx 10^9$ cm^{-3}. The deep donor level is almost free from electrons. The concentration of the equilibrium holes $p_0 = n_i^2/n_0 \approx 10^6$ cm^{-3}.

The holes and electrons are carried out of the SCR by the field; however, due to the thermal transition of the electrons out of the valent zone through the deep level to the conductivity zone, a generation leakage current j_g arises. In addition, a flow of minorities electrons determined by diffusion goes from the NR to the SCR. The density of the flow of holes from the NR of the n layer $j_p \approx qD_p p_0/L_d$ where L_d is the smallest of the two values, either the diffusion length or the thickness of the NR. Since the p_0 value is also determined by the thermal throw, both of these components of the leakage current sharply (in proportion to $\exp(-qE_g/(kT))$ drop with a decrease in the temperature, and, in addition, as was already noted, they are small, and even at room temperature they do not play a part in the initiation of the breakdown.

In the electrical field for the accepted system of levels (fig.5.4), the following types of transitions become possible:

1. Tunnel transitions: valent zone -- conductivity zone (T_{vc}); valent zone -- deep level (T_{vr}); deep level -- conductivity zone (T_{rc}). As is well known, the probability of a tunnel transition is proportional to $\exp(-b_T/E)$, where $b_T = 4(2m^*)^{1/2}\Delta E^{3/2}/(3qh)$, m^* is the effective mass of the electron, and ΔE is the height of the barrier. In our case: $\Delta E = E_g$ for the zone -- zone transition; $\Delta E = E_g - E_r$ and $\Delta E = E_r$ for the transition from zone to deep level and from the level to the zone. For a zone -- zone transition in silicon, as was noted, $b_T \approx 4 \cdot 10^7$ V/cm, and the generation rate in fields with $E < 10^6$ V/cm is very small. Therefore, in diodes with a stationary breakdown voltage of more than tens of volts, the tunnel effect is negligibly small as compared with the impact ionization .

Tunnel transitions with the "participation" of a deep level (we will call them level transitions) have a significantly smaller barrier height and therefore have an exponent which is larger by many orders of magnitude. However, the pre-exponential factor is proportional to the concentration of the levels and can be small. The role of such transitions in the breakdown of high-voltage diodes has not been studied up to the present time, but it is common knowledge that in tunnel diode, it is precisely due to these transitions that the excess currents arise]. For the shallow donor level, these very types of tunnel transitions are very probable. Impact ionization transitions: zone -- zone (U_{vc}); valent zone -- deep level (U_{vr}); deep level -- conductivity zone (U_{rc}), since both electrons (shown in fig. 5.4) and holes (not shown) can cause impact ionization. Ballistic electrons (avoiding collisions with an energy loss on the entire path of the ionization energy collection) cause ionization in not very strong.fields . The probability of such transitions is proportional to $\exp(-b/E)$, where $b = \Delta E_i/(q\lambda)$, λ is the length of the free run without collision with an optical phonon, ΔE_i is the ionization threshold. As was noted above, during zone -- zone ionization, due to the law of momentum conservation, for current carriers the ionization threshold is somewhat higher than the barrier ($\Delta E_i > E_g$). For ionization with level transitions, the pre-exponential factor is proportional to the concentration of the deep levels, and in general will be smaller than for a zone -- zone transition .

For impact ionization, the pre-exponential factor is also proportional to the concentration of the ionizing particles; that is, the ionization rate is proportional to the number of

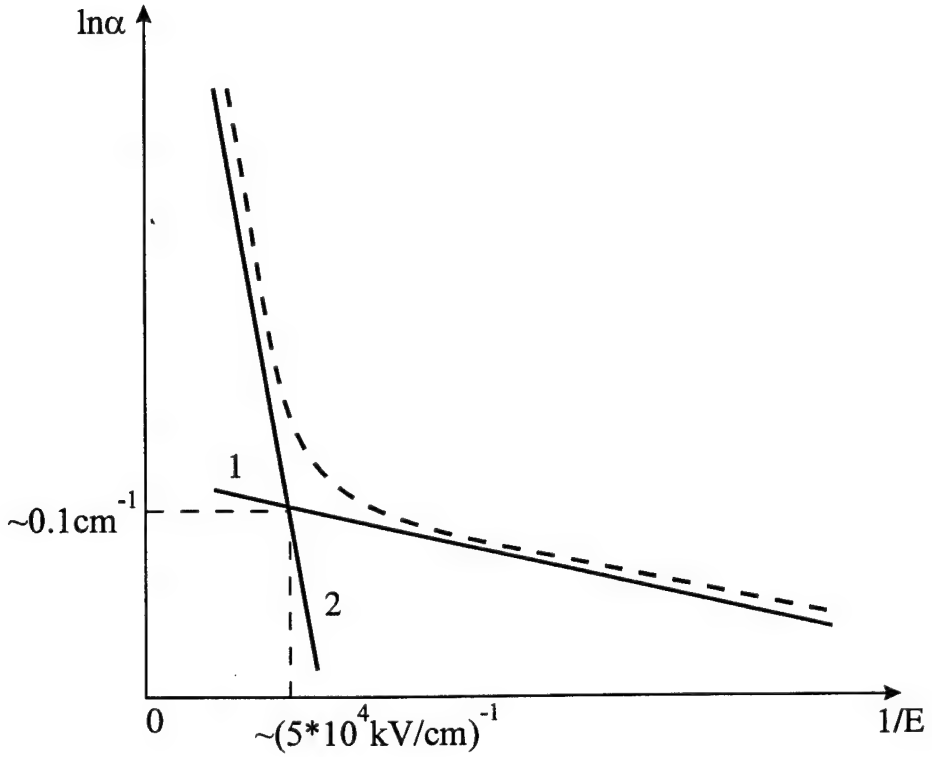


Fig.5.5
The impact ionization coeffitient with level (1) and zone (2) transitions.
The dotted curve denotes the total ionization coeffitient.

carriers. Just as is true for tunnel ionization, there is no precise analytical theory which makes it possible to compute the coefficients for impact ionization in a specific semiconductor while taking the zone structure into consideration. For silicon, this calculation was done by numerical methods [19]. The calculation showed that, taking the upper valley whose position still has not yet been precisely determined into consideration, it is possible to make a noticeable change in the ionization coefficient. There is no reliable information about the structure and parameters of the deep levels, and, consequently, about the impact ionization coefficients as well. The position worsens still by the fact that, like for tunnel ionization, the exponent's extremely high (exponential) sensitivity to error in the indicator, when the exponent has large values, impedes even the evaluation procedure.

Usually, the theoretically obtained parameters are corrected for the best agreement with experimental data. However, for ionization with level transitions, there is almost no such data. Nevertheless, in the simplest case, for a level lying in the middle of the forbidden zone, it is easy to construct a curve for the complete impact ionization coefficient in generally accepted coordinates $\ln\alpha$ and E^{-1} (fig. 5.5). In the case of level ionization, the momentum can be changed with a level; therefore, the threshold ΔE_i will be close to the height of the barrier ($\Delta E_i \approx E_g/2$). In this case, for silicon, the slope of linear functions 1 and 2 will differ by approximately three times. However, due to the large difference in the pre-exponential factors, zone -- zone ionization will play the main part in strong fields, and level ionization will play the main part in weak fields. The summation curve should undergo a sharp break in the area of the cross section of curves 1 and 2. It should be noted that such a break was observed by us [20].

The evaluation for the ionization threshold in the assumption of the unchangeability of λ yields $\Delta E_i = 0.5$ eV, a value which corresponds to the middle of the forbidden zone [20]. Just as is true for thermal ionization, it should be considered that the maximum contribution to the generation of electron-hole pairs should be made by the levels which lie in the middle of the zone. In other cases, the ionization rate sill be determined by the largest energy gap.

When there is a more complicated system of levels, for example when excited states exist, cascade ionization with a still smaller character value for the field intensity in the exponent indicator, but with a smaller pre-exponential factor, is possible. In this case, a second break on curve $\ln\alpha(E^{-1})$ will be observed in the area of still weaker fields.

Cascade ionization is well known in gases, but was not studied in semiconductors.

As was noted, the generation rate in the case of impact ionization is proportional to the number of carriers; therefore, when they are absent, it can come down to zero, and tunnel ionization can be the sole source of carriers. When there is a given carrier flow, it is possible to draw a general conclusion about the relationship of the impact and tunnel components of the generation rate. For the tunnel process, the exponent indicator is proportional to $\Delta E^{3/2}$, while for the impact process, it is proportional to ΔE; therefore, the fraction of the tunnel process is small when there are large ΔE values; however, with a decrease in ΔE, it should increase.

Actually, as was noted above, in the case of a stationary breakdown in silicon, tunnel ionization is unnoticeable on the background of impact ionization. As follows from experiments on the microwave ionization [20], when the concentration of the initiating carriers $n_0 \approx 10^{14}$ cm^{-3}, the impact ionization rate in fields with an intensity of up to 10^5 V/cm in an NR significantly exceeded the threshold ionization rate in the SCR, where the field intensity was almost two times greater. At the very same time, when $n_0 \approx 4 \cdot 10^{13}$ cm^{-3} for ionization with a shallow phosphorus level in a silicon structure ($\Delta E = 0.045$ eV), the tunnel processes are decisive factors even in fields with intensity $(1 - 2)\times 10^4$ V/cm .

Generally speaking, mixed tunnel-impact ionization processes, which decrease the ionization threshold, are also possible]. The effect of tunnel ionization increases as the transition to the strong field region takes place, gradually becoming a decisive factor. However, there are almost no theoretical and experimental data on such mixed processes.

Let us examine in more detail the contribution of the possible sources of initiation carriers for the case which interests us -- that of a p^+-- n -- n^+ structure with a sharply nonhomogeneous distribution along the current lines for both the field (fig. 5.3) and the carriers as well. As was noted, when a constant voltage is applied, in the NR ($\chi > a$), equilibrium carriers are present (n_0, p_0), the deep level is filled by electrons , the donor level is partially full [see formula (5.22)] (fig. 5.3). When increasing the voltage, the SCR expands with a rate $da/dt = v_a$, and in the NR there arises a field with intensity E_{NR}. In the expansion area of the SCR ($a_0 < \chi < a$), ionization of the levels begins. Even when there is an intensity of $2 \cdot 10^4$ V/cm, the time for tunnel ionization from a shallow phosphorus layer is less than 10^{-10} s . Such a field intensity is reached at a distance of approximately $2 \cdot 10^{-3}$ μ from the boundary between the NR and the SCR. Therefore, even when the size of the SCR is approximately 10^{-2} cm in its biggest part the donor levels are completely free of electrons. The electrons thrown out of the levels are carried to the side of the weaker field, which is in the NR.

The equilibrium holes are carried out by the field from the NR to the SCR; however, cooling can completely eliminate this flow. At the very same time, due to the large concentration of the major carriers, the impact ionization rate in the NR can turn out to be rather high, in spite of the field intensity which is small in comparison with the SCR. Such a situation was realized in experiments on microwave ionization [20].

The holes generated in NR due to impact ionization are also carried out by the field into the SCR. The complete flow of holes, as it moves through the SCR to the point of the maximum field, continues to increase due to the impact ionization, and finally a fast ionization wave going in the reverse direction (from the p^+ layer to the n^+ layer) is generated.

Let us evaluate the possibility of tunnel ionization in the area of the maximum field in the SCR. Supposing that tunnel ionization can go through the very same intermediate level as impact ionization can, we get for silicon a b_T value which is $2^{3/2}$ times smaller than zone-zone tunneling; that is, $b_T = 1.4 \times 10^7$ V/cm. In fields realized experimentally with a maximum intensity of approximately $4 \cdot 10^5$ V/cm, zone -- zone tunnel ionization was extremely improbable, but a decrease in the threshold by $2^{3/2}$ times increased the probability by almost 10^{28} times(!). However, although due to the reasons noted previously it is impossible to evaluate the rate of tunnel ionization through the level, some experimental data which will be discussed later (in particular, that for a super-fast increase in the voltage) bear witness to the possibility of such a process.

The carriers which came to be by means of tunnel ionization in the area of the field maximum initiate intensive impact ionization, which can also lead to the excitation of a fast ionization wave according to the case described previously.

One can see, that the examined sequence for the processes (the primary wave of holes from the NR, the fast ionization wave) is not universal. As was noted, in the presence of a certain system of deep levels, the tunnel ionization through it can lead directly to the excitation of the wave in the area of the field maximum in the SCR. Nevertheless, within sufficiently wide limits for the change in U' , processes are observed which take place in the calculated sequence. The experiments will be described below.

In conclusion, one should make a judgment about the possibility of the photogeneration of carriers in front of a wave caused by radiation recombination in concentrated plasma behind the wave front. It was shown, that such a mechanism, for example, in silicon due to

the smallness of the radiation recombination coefficient, can have an effect only in very thick structures ($w \geq 1$ cm) . For the silicon structures with a thickness of approximately 10^{-2} cm, which were discussed in this paper, the effect of the radiation is negligibly small. For direct zone GaAs semiconductors, radiation can even play a substantial role in the process of the generation of the ionization wave.

5.2.4 Impact ionization in the neutral region.

The distribution of holes in the NR in the case of the condition $E_{NR} > E_s$ is described by the continuity equation

$$\frac{\partial p}{\partial t} + v_s \frac{\partial p}{\partial \chi} = G. \qquad (5.23)(5.23)$$

Ionization in the NR is initiated by a large quantity of electrons. Since in relatively weak fields, the number of secondary carriers which came to be due to ionization during a small time interval is a lot smaller than the number of primary carriers, it is possible to consider the generation function to be constant with respect to coordinate, and to write it in the following form: $G = N_d v_s \alpha$. The solution to equation (5.23) with zero (initial and boundary conditions on the n^+-- n transition) conditions is recorded in the form:

$$p = v_s N_d \left[\int_0^t \alpha dt - \int_0^t \alpha\left(t - \frac{\chi}{v_s}\right) F\left(t - \frac{\chi}{v_s}\right) dt \right], \qquad (5.24)$$

where

$F\left(t\frac{-\chi}{v_s}\right) = \begin{cases} 0 \text{ when } t < \chi/v_s; \\ 1 \text{ when } t > \chi/v_s \end{cases}$

is the step Heavyside function.

When solving equation (5.23), the beginning of the $\chi = 0$ coordinates can be conveniently placed at the n -- n^+ transition. Expression (5.24) means that for each moment in time t, the concentration p does not depend on the coordinate when $\chi > tv_s$:

$$p = N_d v_s \int_0^t \alpha dt. \qquad (5.25)$$

At the boundary of the NR with the SCR ($\chi = a$) which is shifting toward the beginning of the coordinates with speed v_s(that is, $a = a_s - v_s t$, where a_s is the position of the boundary up to the moment that it reaches saturation speed $t = \tau_0$), the concentration of the holes will be

$$p = N_d v_s \left[\int_0^t \alpha dt - \int_0^{t-\frac{a_s}{v_s}} \alpha\left(t - \frac{a_s}{v_s}\right) dt \right] \qquad (5.26)$$

where $t > a_s/(2v_s)$, when $t < a_s/2v_s$, the expression for p has an appearance which is similar to (5.25).

In order to make a calculation according to formula (5.25), we use the usual expression $\alpha = \alpha_\infty e^{-b/E_{NR}}$. Since the field and the time are connected by dependence (5.17), the replacement of the variable brings expression (5.25) down to the following:

$$p = G_0 \int_{E_s}^{E_{NR}} \frac{e^{-b/E_{NR}}}{\sqrt{4d}\sqrt{E_{NRm} - E_{NR}}} dE_{NR}, \qquad (5.27)$$

where $G_0 = N_d v_s \alpha_\infty$; $d = U_k' v_s/(2a_s w)$.

It is possible to divide the integrating area into two. In the area of values $E_{NR} < E^*$, the exponential part changes the most quickly, and in the area where $E_{NR} \geq E^*$ the factor $(\sqrt{E_{NRm} - E_{NR}})^{-1}$ which strives toward infinity when $E_{NRm} \rightarrow E_{NR}$ changes the most quickly. In this case, taking the slowly changing cofactor out for the integral and integrating, it is easy to get $p = p_1 + p_2$, where

$$p_1 = \frac{G_0 E_{NR}^2 \exp(-b/E_1^*)}{2b\sqrt{d}\sqrt{E_{NRm}-E_1^*}} \text{ when } E_{NR} \le E^*; \quad (5.28)$$

$$p_2 = \frac{G_0 \exp(-b/E_2^*)\sqrt{E_{NR}-E_2^*}}{\sqrt{d}} \text{ when } E_{NR} \ge E^*. \quad (5.29)$$

It is possible to find the E^* value when equating values p_1 and p_2 at the overlapping point: $E_1^* = E_2^* = E^*$, which yields

$$E^* = E_{NRm}\left(1 - \frac{E_{NRm}}{2b}\right). \quad (5.30)$$

It is easy to show that for the time that a field with intensity $E_{HO} > E_s$ exists, the concentration increases to the maximum: $p_m = 4p_2$.

Using the values $\alpha_\infty = 2.6\ cm^{-1}$ and $b = 2.8 \cdot 10^5$ V/cm in weak fields [20], from expressions (5.17), (5.29), and (5.30) it is easy to obtain the evaluation for p_m. When $N_d = 10^{14}$ cm^{-3}, $U_0 = 400$ V, we get $E_{NRM} = 1.8 \cdot 10^4$ V/cm, $p_m = 10^5\ cm^{-3}$ for $U' = 2 \cdot 10^{12}$ V/s and $E_{NRm} = 6 \cdot 10^4$ V/cm, $p_m = 10^{10}\ cm^{-3}$ for $U' = 3 \cdot 10^{12}$ V/s.

Thus, the unequlibrium concentration of holes which came to be due to ionization in the NR can substantially surpass the equilibrium concentration. Let us emphasize that the obtained evaluation is very crude and is not general in character. In reality, ionization through the system of intermediate levels can have an extremely strong dependence on the type and concentration of uncontrollable levels contained in a specific model.

We should explain the remark which was made in more detail.

The position here is evidently close to that which was created in the question of the leakage current in high voltage p -- n junctions in silicon. It is acceptable to think that this current is determined by thermal generation in the SCR through the deep level (or system of levels). In agreement with the Shockley -- Sa -- Noice theory, the levels which lie close to the middle of the forbidden zone make the greatest contribution. In the opposite case, the speed of the transition of an electron from the valent zone to the conductivity zone is small, since it will be determined by a small probability of thermal ionization through the greatest energy gap. When introducing deep levels (controllable) in a sufficiently large quantity, for example when doping using gold and other impurities, it was reported that the thermal generation through these levels really determines the leakage current in agreement with the Shockley -- Sa -- Noice theory. However, even in "pure" p -- n junctions where the presence of deep levels fails to be detected by independent methods, the leakage current significantly surpasses the calculated diffusion current from the neutral regions and the current of the thermal direct zone -- zone generation. A similar position is also the case with recombination, when, in a very "pure" material, the lifetime of the minority carriers all the same turns out to be many times lower that the "radiation" time. In all of these cases, the existence of thermal ionization (or recombination) processes through unidentifiable deep levels is assumed

The sources of primary carriers in an SCR which are "switched on" by an electrical field, which arises during a sharp increase in the voltage on the diode, were discussed above. It can be expected that these very unidentifiable deep levels, which determine the thermal generation (and, as a rule, recombination as well), also determine the field generation. In the case of impact ionization through deep levels, the threshold values for the ionization coefficient $\alpha_{\infty r}$ are small -- a lot smaller than during zone -- zone ionization. Let us recall that for impact ionization through levels, that very law $\alpha = \alpha_{\infty r} e^{-b/E}$, which is also for the zone -- zone process, the ionization integral $J = \int_0^{w_{SCR}} \alpha d\chi < \int \alpha_{\infty r} d\chi$, determining the amplification of the leakage current of the p -- n junction $M = 1/(1 - J)$, is always small. So, $J < 10^{-2}$ ($M < 1.01$) when $\alpha_{\infty r} \approx 1 cm^{-1}$ and $w_{SCR} = 10^{-2}$ cm. Thus, the effect of impact ionization through a level

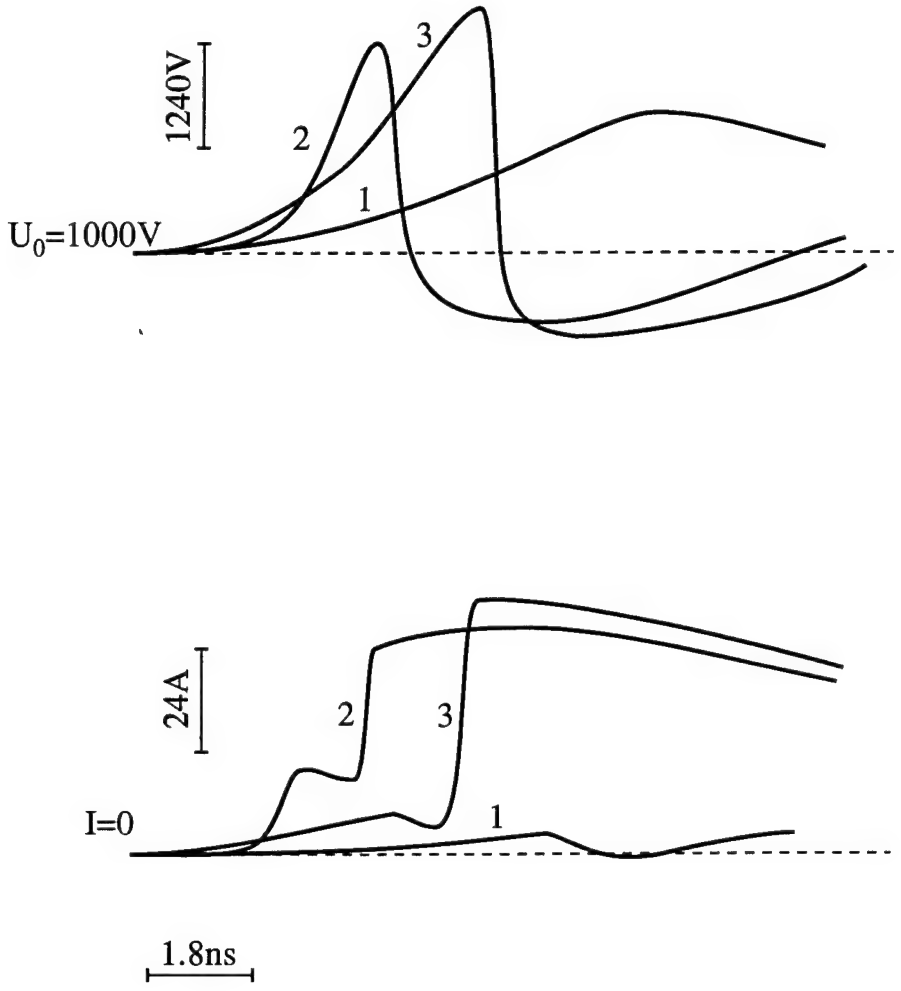


Fig.5.6
Switching on of the diode, when different U' is applied.

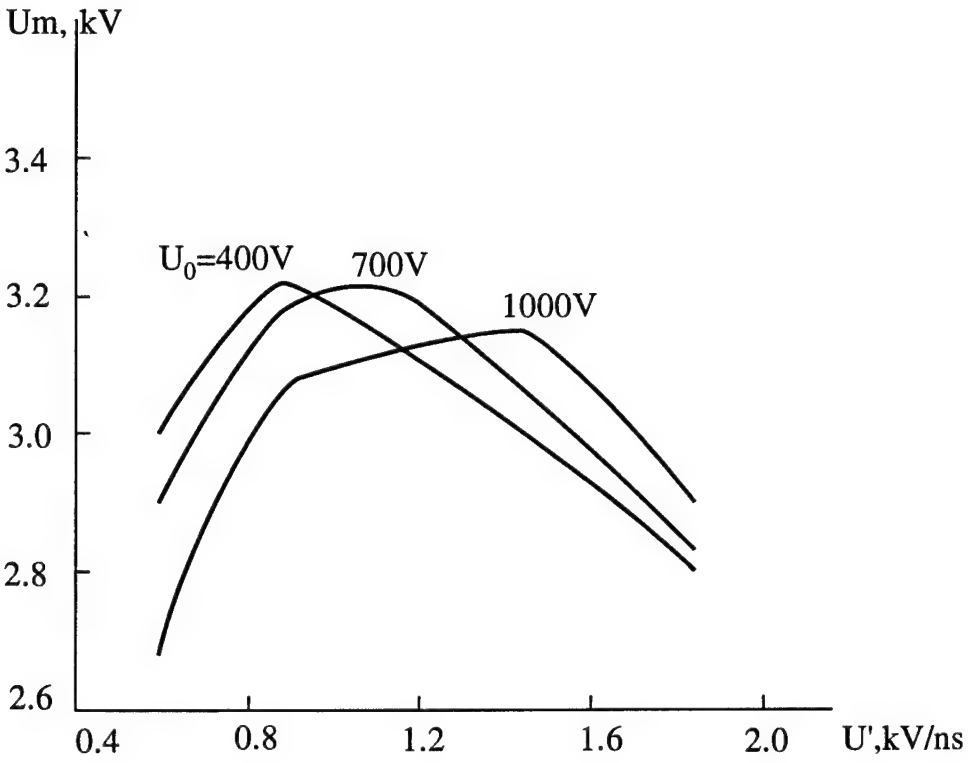


Fig.5.7
Turn on voltage versus applied voltage rise rate.

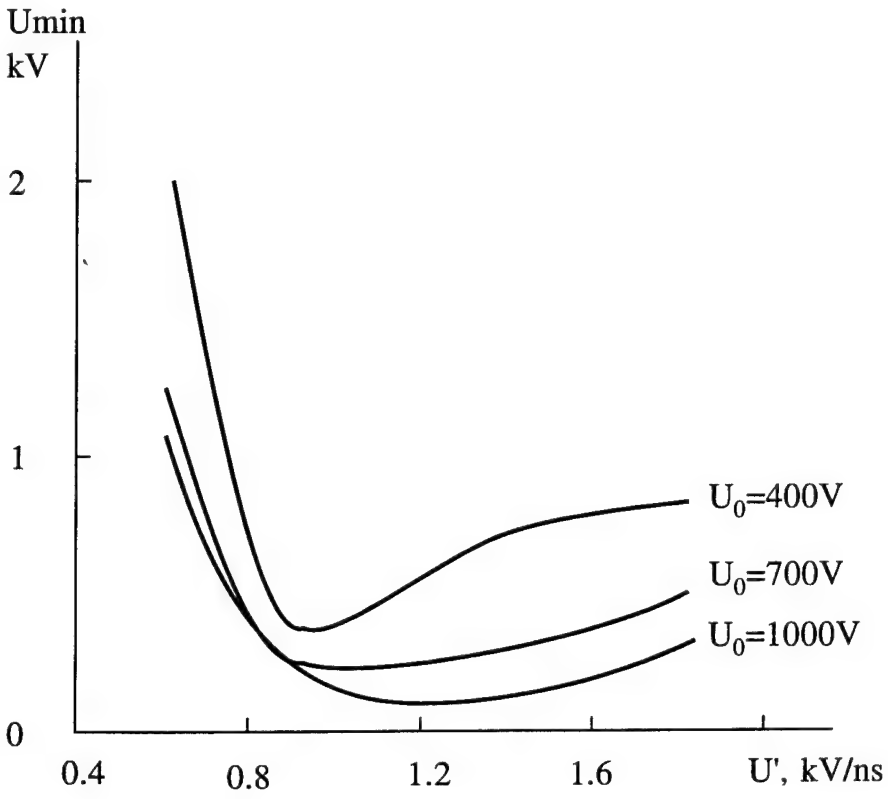


Fig.5 .8
Residual voltage versus applied voltage rise rate

in the SCR is always negligibly small. At the very same time, in the NR this process can play a substantial and even decisive part.

The presence of deep layers, as was already noted, could significantly increase the probability of tunnel ionization in the SCR. Experimental confirmations of this will be given below.

5.3 The effect of the parameters of the system and the structure on the super-fast switching process

5.3.1 The effect of the circuit parameters

At fig.5.6 there are shown the results of experiments on the study of the effect of U' and the constant bias voltage U_0 on the process of switching p^+-- n -- n^+ structures with: $N_d = 10^{14}$ cm^3; w = 250 μm.. The U' value decreased by use of low pass LC filters with the corresponding frequency cut off

It is evident that the change in U' has a very strong effect on the character of the switching. When U' < 0.5 · 10^{12} V/s, the effect of the switching is practically absent. When there are larger U' values, a "precursor", determined by the displacement current, grows in front of the section of the rapid increase in the current. Experiments shows that, when there is a small constant bias and U' = 10^{12} V/s, there is no switching, and when there is an increase in the bias more than 400V the switching on appears. It should be noted that when there is a bias voltage close to the stationary breakdown voltage, the effectiveness of the switching decreases. This is connected with an increase in the leakage current and it is similar to the behavior of a structure when heated or illuminated.

As is evident from fig. 5.7, the dependence of the maximum switching voltage on the rate U' could be bell-shaped in character. When there are small rate, it increases with an increase in U', and when there are large rates it drops. Moreover, when there are small U' values, the increase in the bias voltage U_0 leads to a decrease in U_m, and when there are large U' values, it leads to an increase in U_m.

The dependence of the residual switching voltage after switching on U' has a minimum (fig. 5.8): when there are small rate (U' < 10^{12} V/s), an increase in U' leads to a sharp decrease in U_{min}, and when there are large rate (U' > 10^{12} V/s), it leads to certain increase.

The dependence of the charge carried from the diode at the plasma extraction stage P_e and of the equal charge created by the ionization wave on U' also is bell-shaped, which is the opposite of the dependence of U_{min} on U' . When there are small speeds, a sharp increase in P_e with an increase in U' is observed, and when there are large speeds, a certain drop is observed

In order to interpret the experimental results, let us look at the dynamics of the electrical field before switching. In section 5.2, an approach was developed which makes it possible to calculate the field in a structure according to the given rate of growth of the applied voltage. There was shown that in NR the field dependence on time has a bell-like shape and the increase of bias voltage (and so of SCR width) leads to decrease of field maximum value.

When U' < 0.5·10^{12} V/s, the E_{NR} value is very small, and ionization in the NR is lacking. The equilibrium holes thrown out of the NR to the SCR go to the maximum of the field in approximately 1 ns ($a_0 \approx 10^{-2}$ cm when U_0 = 1 kV), when the voltage on the structure only reaches the voltage of the stationary breakdown 1.5 kV. Subsequently, the voltage will increase under conditions of a large enough flow of holes ($p_0 \approx 10^6$ cm^{-3}); that is, stationary breakdown begins without switching (without an ionization wave).

A similar situation also arises in that case when the initial constant voltage bias is very small . When U_0 = 0, the dimension a_0 = 0 and the holes from the NR almost

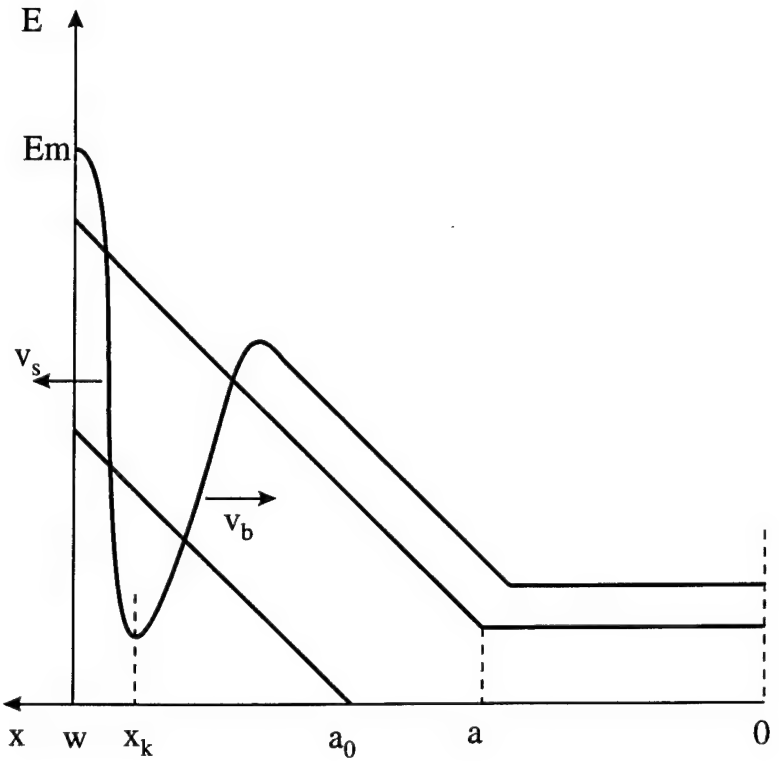


Fig.5.9
The redistribution of the field during the formation of an ionization wave in the middle of SCR.

Premature wave exitation.

immediately get into in the region of the field maximum; that is, an increase in the voltage higher than the threshold of the stationary breakdown occurs when there is a large concentration of initiating carriers and switching is also absent. It may be shown, when there are super-high U' values, a large initial concentration does not obstruct the formation of waves if criteria $\frac{2bqp_0}{E^2 \varepsilon\alpha} < 1$ (where p_0 is initiating carriers concentration, b is characteristic field) is fulfilled. When there are large U' (greater than $2 \cdot 10^{12}$ V/s), the initial flow of the holes which came to be in the NR by means of ionization has a large density. The field intensity in the SCR will increase faster; therefore, the rate of growth of the carrier concentration in the front of the hole packet which entered the SCR also increases. Thus, the concentration in the front reaches the critical value N_d earlier, and at this critical value there begins the distortion of the distribution of the field by means of the space charge of the mobile carriers; that is, the ionization wave arises closer to the boundary between the SCR and the NR, which is at critical point χ_k (see fig. 5.9), positioned far from the field maximum. The concentration of the plasma behind the front of the wave [see expression (5.3] strongly depends on the field in front of the front; therefore, the wave formed "beforehand" poorly modulates the volume of the semiconductor, and therefore the residual voltage is great and the charge at the base is small. It is obvious that the optimum condition for switching (the maximum switching voltage and the minimum residual voltage) is when a fast wave is formed at the point of the field maximum, to which corresponds the definite ratio between the rate of growth of the voltage, the density of the flow of holes, and the initial dimension a_0. The existence of such an optimum is clearly seen from fig. 5.7 - 5.8

We can determine from the current traces the pulled out charge P_b and, assuming that the distribution of the concentration with respect to the volume after the wave run is homogeneous, the average concentration. In case of optimum switching we have got ($U' = 1.2 \cdot 10^{12}$ V/s, maximum P_b): $n_m \approx P_b/(qSw) \approx 10^{15}$ cm^{-3}.

In the very same supposition, we can find n_m , knowing the residual voltage U_{min}(see fig.5.8): $n_m \approx U_{min}/(q\mu jw) = 10^{15}$cm^{-3}.

Before switching the diode current is displacement current, therefore we can get the maximum of the field intensity near P-n junction as

$$E_m = E_0 + \frac{1}{\varepsilon S}\int_0^t Idt \qquad (5.30)$$

According to(5.30), it is easy to find the value $E_m = 3.3 \cdot 10^5$ V/cm for the moment of the time just before the switching for which the evaluation of the concentration according to formula (5.3) yields $n_m = 5 \cdot 10^{15}$ cm^{-3}.

During the total switching time $\tau_\varphi = 0.2$ ns, the wave should pass through the SCR and reach the NR, which from the time that the switching begins succeeds in decreasing by no more than $\Delta a = 2 \cdot 10^{-3}$ cm; that is, to width $w - a_m - \Delta a = 7 \cdot 10^{-3}$ cm. If the field intensity in front of the wave front retained the E_m value, then the voltage drop in the remaining portion of the NR would be no less than 2 kV. It is clear that, due to the decrease in the voltage on the structure as the current increased, the field in front of the front should weaken in comparison with the initial one, and, consequently, the concentration n_m should decrease. Taking what has been stated into consideration, and also the sensitivity of the n_m value to the change in E_m [in agreement with formula (5.3] and the inaccuracy of the determination of S and U_{min}, the coincidence of the three values for n_m (determined by independent methods) should be considered completely satisfactory. This coincidence also indicates the sufficient homogeneity of the modulation process. Since expression (5.3) yields an evaluation from above, then from a comparison of this evaluation an the evaluation of the concentration according to the carried out charge, it follows that modulation enveloped no less than 40% of the area of the device.

The experiments during which the area of similar structures was gradually decreased showed that the residual voltage increases with a decrease in the area according to a law which is close to the linear law. Therefore, it is possible to confirm that in structures with a large NR thickness, the wave process is developed homogeneously enough. This homogeneity, which is caused by the large concentration of the carriers initiating the wave, also explains the extreme stability of the process in time .

Thus, the given series of experimental data bear witness to the benefit of the fact that in "thick" structures, which are those with a large NR and a large initial SCR when there are moderate U' values, the process develops according to the system examined above: injection of carriers from the NR to the SCR, and excitation of a fast ionization wave.

5.3.2 . The effect of the structure's parameters

The problem of the effect of the technological parameters of a structure such as the degree of doping, the thickness, and the depth of the bedding of the junctions on super-fast switching in a wide range of systemic parameters U_0, U', and R, is very interesting and important. However, such studies are extremely difficult. An extremely wide range of such characteristics which are specific for the discussed phenomenon as the maximum switching voltage, the residual switching voltage, and the duration of the switching process can be observed from structure to structure. The laws of the dependencies (U_m on U_0, etc.), can even change qualitatively. Within a batch manufactured in one technological cycle, the spread is usually very small. An uncontrollable spread in the parameters arises between structures which are manufactured in different technological cycles, out of a different initial material, although such parameters as the depth of diffusion, the thickness of the structure, the surface and volume resistance, the carrier lifetime, and the voltage of the stationary breakdown are controlled well enough. Such a position is evidently explained by two causes.

1. As was noted previously, the wave ionization processes are closely connected with the processes for the formation of primary initiating carriers, since in the super-voltage region, weak tunnel ionization through the system of deep levels in the absence of impact ionization can be a determining factor. The very same thing applies to impact ionization through the level in the NR, where zone -- zone ionization impossible due to a weak field. The control of field ionization in our case is extremely difficult, just as is true of the control of thermal ionization through deep levels in the SCR (in silicon high-voltage device) which determine the leakage current of a p -- n junction. In spite of all attempts to purify the material from all levels lying by the middle of the forbidden zone, and making the main contribution to the leakage current, residual uncontrollable levels (or a system of levels), which also guarantee an uncontrolled leakage current which is greater than the calculated diffusion currents from the neutral regions, always exist. The concentration of these levels can be so small that it is impossible to determine it in other independent experiments (the measurement of photo- and thermocapacitance, etc.).

2. Generally speaking, ionization waves can be unstable with respect to spatial perturbation with a wavelength which is greater than the thickness of the front of the ionization wave (see section 5.1.1), since the latter, with an increase in speed, becomes more unstable. In a semiconductor, the development of instability which is manifested as the current's "filamentation" effect, leads to the breakdown (burning out) of the structure. Actually, in systems of very large super-voltages and small fronts, the switching in thin structures was unreliable, and the structures very quickly (often after the first switching) burned out, of course, if the duration of the pulse was large (10 ns). As practical experience has shown, the presence of the NR in the entire range of working voltages is necessary for reliable operation (the

a)

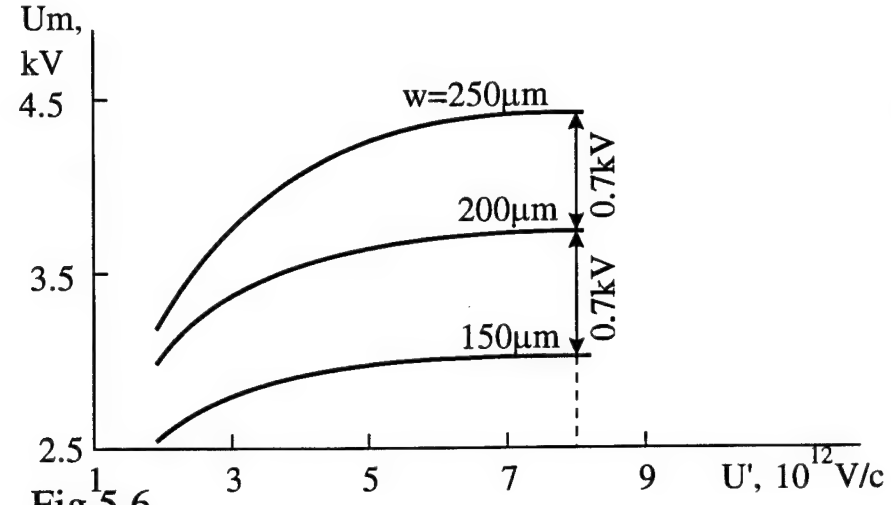


Fig.5.6

b)

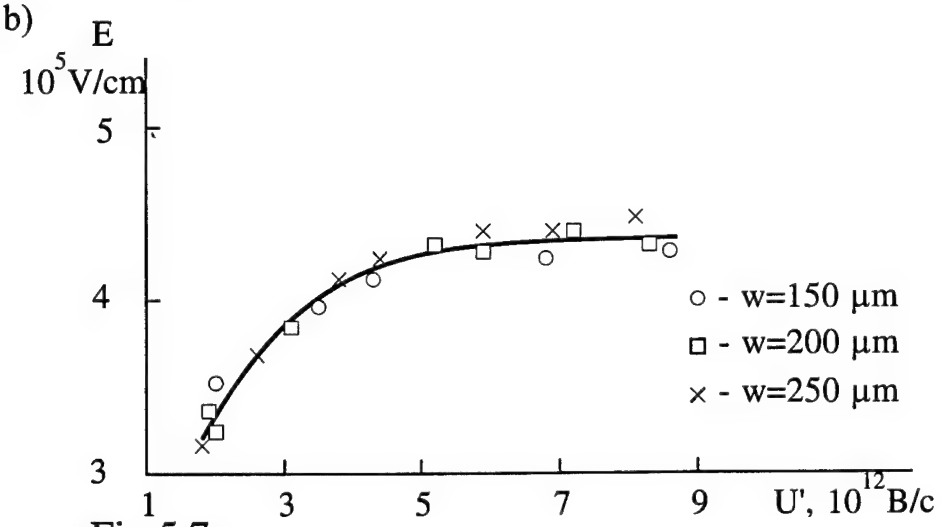


Fig.5.7

c)

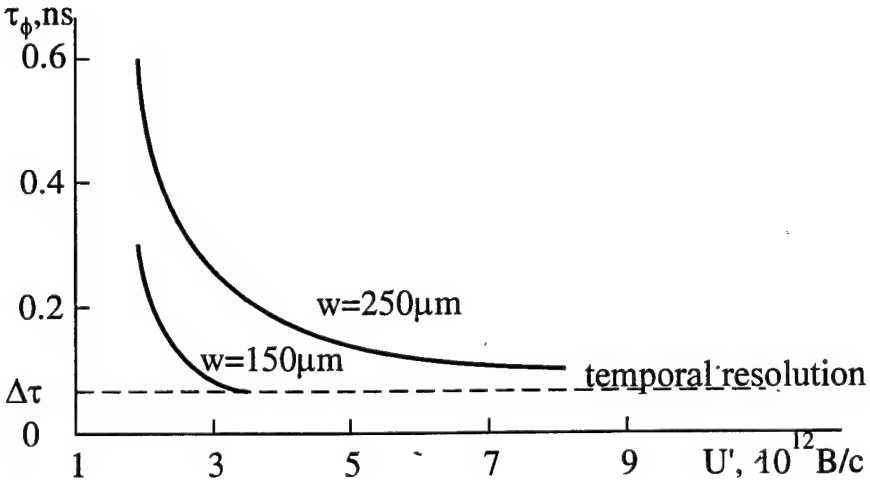


Fig.5.10
The dependances of the switching voltage (a),
maximum field intensity (b)
and duration of the switching (c)
on the rate of increase of the votage on the structure.

overlapping of the base by the space charge area is not permitted). Evidently, the NR, acting as the ballast series resistance for the wave, suppresses instability.

Below are given the experimental data on the effect of the thickness of a high-ohmic n layer in a p^+-- n -- n^+ structure on the switching process. For the reasons stated above, all the structures were manufactured in one technological cycle out of silicon wafers having different thickness and having an identical donor concentration $N_d = 1.25 \cdot 10^{14}$ cm^{-3}. The depth of the doping of the p^+-- n junction which was obtained by means of the diffusion of boron with a surface concentration $N_s \approx 5. 10^{19}$ cm^{-3} and aluminum with $N_s \approx 10^{17}$ cm^{-3}, equaled 50 μm. In fig. 5.10 a,b,c there are given the dependencies of the switching voltage, the maximum field intensity at the p -- n junction at the switching moment [calculation is based on the experimental data according to formula (5.30)], and the turn on time on the rate of increase of the voltage for different thickness of the n layer. As was stated previously, when supplying a constant voltage, the switching of structures with a small thickness of the n layer, and thin NR as well, became very unreliable (the devices quickly burned out). Therefore, all of the data were obtained when there was nearly zero constant bias.

For large U' values (more than $2 \cdot 10^{12}$ V/s), the structures switch on well, and without constant displacement. In this case, the large flow of primary carriers does not prevent the creation of a overvoltaged region, and, as was noted above , fast switching is possible. However, the gain with respect to fast response should decrease sharply, which is also observed in experiments.

As follows from fig. 5.10, the increase in the rate U' at first leads to an increase in the voltage U_m. When $U' > 7 \cdot 10^{12}$ V/s, this dependence is obviously saturated, but there is no previously observed (see fig. 5.7) section where there is a drop in U_m with an increase in U'. We think, that we have not had enough large voltage rise rate to get to the falling part of the curve. When there are larger U' values (10^{13}V/s), the characteristic time for the increase in the voltage $\tau \approx U_m/U' \approx 0.3$ ns. During this time, the SCR will not be able to expand to a size larger than $\Delta a = 30$ μm. Consequently, the jump in the field intensity at the SCR $\Delta E < qN_d\Delta a/\varepsilon \approx 6 \cdot 10^4$ V/cm. The voltage drop which corresponds to this jump is $\Delta U = \Delta E\Delta a/2 = 10^2$ V; that is, it equals a small fraction of the complete voltage on the structure. This means that the field will be distributed homogeneously almost along the entire n layer, with the exception of the SCR with which Δa, and with an intensity $E_{NR} = U_m/w \approx 1.7 \cdot 10^5$ V/cm. Increasing the thickness of the n layer by 50 μm leads to an increase in the voltage by approximately 850 V, which is close enough to the data in fig. 5.10 a (approximately 700 V). The small discrepancy can be explained by the crudeness (underestimation) of the evaluation of Δa, which does not take into consideration the specific course for the increase in the voltage in time, which is with a precursor, in the course of which the SCR expands.

. The calculation of the field intensity at the p -- n junction at the moment of switching while taking into consideration the true course of curve U' (see fig. 5.10 b) shows that with an increase in U', the E_m value at first increases linearly, and then, when $U' > 5 \cdot 10^{12}$ V/cm, it is saturated, striving toward a value of $4.4 \cdot 10^5$ V/cm.

It is evident from fig. 5.10 b that, for all structures with different n layer thickness, the dependencies coincide with good accuracy. It should be assumed that when there are large U' speeds and a small size for the SCR, the switching process will take place along a different route than in the preceding case (see section 5.2.4), which is when there are small U' values and a large SCR. In agreement with expression (5.30), in fields with an intensity of approximately $4 \cdot 10^5$ V/cm, the excitation of a fast ionization wave is possible even when there is a very large carrier concentration: $n_0 \approx E_m^2\varepsilon a/(2qb) \approx 10^{16}$ cm^{-3}. When $U' > 5 \cdot 10^{12}$ V/s in experiments, the current density before switching with a small error of less than 20% is

determined by the law $j = CdU/dt(C = \varepsilon S/w)$; that is, the current is the displacement current, and its density does not exceed 500 A/cm^2. Thus, the upper evaluation for a density of the conductivity current equaling $j_{mp} < 100$ A/cm^2, and for the concentration - $n_0 \leq j_{mp}/(qv_s) = 10^{14}$ cm^{-3}, and this value is a great deal less than the threshold value of 10^{16} cm^{-3} (obtained above), which bears witness to the possibility of the excitation of a fast wave.

A necessary condition for the wave process is the gradient of the field and (or) concentration. There is a field gradient in the SCR; therefore, the ionization process can be a wave process. In the NR, the distribution of the field and the concentration is homogeneous from the start. The breakdown of homogeneity can arise due to the generation of unequlibrium carriers by impact ionization and the increase in their concentration to a value which is comparable with N_d. But even in this case, the nonhomogenity region, positioned at the boundary of the NR by the n^+-- n junction, will have dimensions $\Delta \approx v_s\tau$. Therefore, when there is a small τ, and a small constant bias (and , consequently, $a_0 << w$) after the exit of the wave from the SCR to the NR, the entire voltage will be quickly "thrown" at the NR, and in it there begins homogeneous delayed ionization. It can be shown, that such a process also leads to switching; that is, it leads to an increase in the current when there is a decrease in the voltage, but with a small gain with respect to fast response.

The saturation of the dependence of E_m on U' (when $E_m \approx 4.7 \cdot 10^5$ V/cm) evidently means that intensive tunnel ionization is beginning through the intermediate level in the forbidden zone, which was discussed above, in section 5.2.3.

As follows from fig. 5.10, with an increase in U', the switching time τ_φ decreases, and with an increase in the thickness of the n layer it increases proportionally. When there are large U' values (more than $5 \cdot 10^{12}$ V/cm), a tendency is observed toward the saturation of τ_φ , which occurs in the very same U' region where the saturation of the dependence of U_m and E_m on U' takes place , which signifies that there is a direct connection between the switching speed and the overvoltage.

5.4 . Diode like switches

In the preceding section, the results of the study of the effect of switching connected with delayed impact ionization were given. This effect can be used for creating super-fast-switching devices. Such devices do not have stationary S-type volt ampere characteristics, and are actually used as peakers of initial pulses. For peakers, one of the most important parameters is the ratio of the operation delay τd, to the duration of the front of the formed change in τ_φ , which almost corresponds to the ratio of the speeds of the increase in the voltage until switching and the drop during switching. This ratio (κ) is the gain with respect to fast response, which is obtained using a peaker. In the case of the development of a peaking element, it is necessary to guarantee the maximum κ value.

As was shown above, a sufficiently large gain ($\kappa \approx 10$) when there is good stability and high operation reliability of the diode can be obtained in a two-step switching process. Such switching is accomplished in diodes with a specific resistance of the initial silicon ρ = 30-50 ohm·cm ($N_d = 10^{14}$ cm^{-3}). In them, the effect of the switching disappears when the rate of the increase in the voltage is less than $0.5 \cdot 10^{12}$ V/s. When there is the maximum switching voltage U_m = 3 kV and U' = 10^{12} V/s corresponds to a delay of $\tau_з$ = 3 ns. For the indicated U' and U_m values, the switching time equals τ_φ = 0.2 ns; that is $\kappa \approx 15$. The thickness of the n layer and the area of the device play an important part. Instruments which are intended for operation with small U' speeds should operate with an initial displacement which is close to the stationary breakdown threshold in agreement with this, they have an SCR with a large thickness with an initial value a_0 = 100 μm when U = U_0 and a final value a_m = 200 μm when U =

U_m. In order to suppress the current's "filamentation" effect, that is, in order to guarantee the reliable operation of the device, the insufficient reserve of NR with respect to thickness is necessary, since the NR prevents the development of instabilities. Therefore, the complete thickness of the n layer should equal 250-300 μm, which also limits the minimum switching time by a value of 0.2 ns.

Decreasing the area of the device decreases the current passing through it at the delay stage, and leads to an increase in the density of the current and the residual voltage after switching. The optimum area when $U' \approx 10^{12}$ V/s and device operats in circuits with an impedance of 50 ohm equals $S = 0.10\text{-}0.15$ cm^2. Such a value guarantees the smallness of the precursor U_C in the shaped pulse (due to the flow of the capacitance current) of less than 15% and the residual voltage $U_{min} \approx 200\text{-}500$ V.

In order to shape pulses in the picosecond range, a large overvoltage and $u' = (3\text{-}5) \cdot 10^{12}$ V/s are necessary. In this case, operation without a constant bias is possible. Since the maximum thickness of the SCR in this mode decreases, then it is possible to decrease the total thickness of the n layer to 200 μm and, having retained the size of NR, to avoid the "filamentation" of the current.

In order to decrease the parasitic capacitance current, it is necessary to decrease the area of the diode down to 0.015 cm^2. This makes it possible to retain roughly those very U_C and U_{min} values which are in diodes intended for operation with small U' (10^{12} V/s). The experiment with such an optimized diode ($w_n = 200$ μm, $S = 0.023$ cm^2) showed the switching voltage 3.3 kV, and the residual voltage is close to zero. The duration of the front of a pulse shaped at a load of 50 ohm, equaled less than 50 ps, and was located behind the threshold for the resolution of the system. The amplitude of the fast change in the load is 2 kV, and the power is 80 kW, which is four orders of magnitude greater than the power commutated by the known devices in the given temporal range. The instability of the switching, which is determined with respect to the jitter of the samplescope, did not exceed 0.02 ns; that is, the instability of the oscillograph itself.

The threshold operating frequency of a diode peaker depends on two factors: the time of the return to the initial state, and heating up. It should be noted that process of plasma extraction after switching on in SAS is just the same as for DSRD. As was shown in section 4 the plasma remaining after switching could be extracted by the current in a small enough amount of time (10^{-8} - 10^{-8}s) and the limit for operating frequency may be as high as several megahertz. However, after the break in the current, in the NR there can remain unequlibrium carriers whose departure can be determined by the diffusion through the NR to the SCR and by the recombination in it. The diffusion flow of these carriers from the NR to the SCR is equivalent to the leakage current and can prevent the creation of a overvoltage region when there are small U' values.

The experiments conducted on two-pulse methodology showed that the restoration time when $U' \approx 10^{12}$ V/s is approximately 2 μs. When increasing the rate of increase of the voltage, the effect of the residual charge in the NR decreases.

The heat losses can conditionally be divided into three parts: loses on the precursor of the current (at the delay stage), at the front, and at the plasma extraction stage. At the delay stage, the current in a significant portion of the structure (SCR) is a displacement current, and is not accompanied by heat dissipation. The energy lost in the NR can be evaluated (from above) in the following manner

$$P_3 \approx j_s E_s S \tau_\delta (w - a0) \approx 10^{-6} \text{ J}.$$

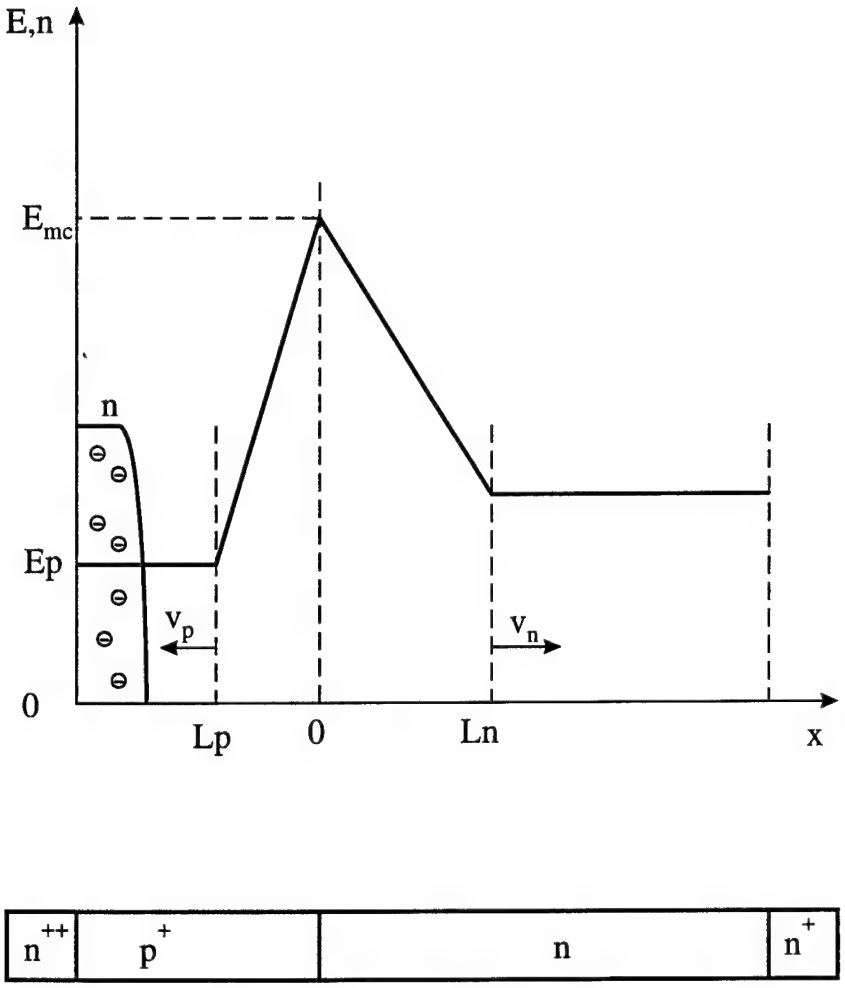


Fig.5.11
Transistor with delayed ionization.

The losses at the front should be less than the usual commutation losses in switches in the approximation of the linear increase in the current, since in the switching process, in front of the wave front, the current is a displacement current. Consequently,

$$P_f \leq \frac{I_m U_m}{6}\tau_f \approx 3\cdot 10^{-6} J.$$

At the restoration stage, it is possible to distinguish two stages: at the first (τ_1) the concentration of the electrons near the boundary with the p^+ layer is great, there is no SCR, and the voltage drop is equal to the residual voltage; at the second stage (τ_2), the restoration of the SCR and the increase in the voltage on the diode begin. The losses on it depend on the specific shape of the pulse sent to the diode. When there is a short pulse ($\tau_p < \tau_1$), the losses are determined by the residual voltage:

$$P_l = \frac{I_m U_{\min}\tau_p}{2} \approx 3\cdot 10^{-6} J$$

When there is a large pulse duration, the losses in the second section can exceed the losses on the first by an order of magnitude. Thus, when there is reliable heat exchange (10 W/cm^2) and a permissible temperature for heating the structure 100^0C, the threshold average operating frequency limited by heat liberation can exceed 100 kHz. Transistor like switches

As was previously noted, after filling the n base of a p^+-- n -- n^+ structure with plasma, as a result of the run of the impact ionization wave, the process of the dispersion of this plasma begins. After the conclusion of the process (after approximately 10 ns), the blocking capability of the p -- n junction is restored. However, in some applications it is necessary to keep conducting state of a switch for a prolonged time interval (10^{-8} s and more).

It is well known that in transistor structures (n^{++}-- p^+-- n -- n^+), the complete charge which passed through the external circuit, due to the regenerative feedback, can significantly exceed the charge preliminary to that "pumped" to the n layer. This is the principle, making it possible to slow down the dispersal of the plasma, which can also be used in devices on impact ionization waves, exciting the wave in the collector p^+-- n -- n^+ part of the transistor structure n^{++}-- p^+-- n -- n^+.

Let us examine the process of switching such a transistor structure (fig. 5.11) .The transistor is connected into a circuit in the same a manner as SAS.

As the applied voltage increases, the SCR near the p^+-- n junction will expand, and in this case the maximum field intensity on the p^+-- n junction increases and reaches the critical value E_{mK}, at which a impact ionization wave is formed. The expansion of the SCR is associated with the removal of the main carriers from the p^+-- n junction with a drift speed in the electrical field having intensity E_p, which arose in the NR of the p^+ layer, equaling $v_{pp} = \mu_p E_p$. This very field causes the injection of electrons from the n^+ layer, and their drift in the direction toward the p^+-- n junction with a speed of $v_{np} = \mu_n E_p$. It is obvious that the electrons go away from the n^{++} layer to a distance $L_n = bL_p$, where $L_p = \varepsilon E_{mc}/(qN_a)$, and N_a is the concentration of the acceptor doping impurity in the p^+ layer.

Thus, when fulfilling the condition

$$w_p \geq (1 + b)L_p, \qquad (5.31)$$

where $L_{pc} = \varepsilon E_{mc}/(qN_a)$, and w_p is the thickness of the p^+ layer, the electrons from the n^+ layer will not be able to end up in the strong field area earlier than the intensity in it reaches the E_{mc} value.

It is obvious that in this case, a wave in a transistor structure can be excited in exactly the same way as in a diode structure. In an asymmetrical p^+--n junction, $N_a >> N_d$ and $E_{mc} = \sqrt{2qN_d U_m/\varepsilon}$, where U_m is the voltage of the switching of the $p^=$-n structure.

After the run of the wave and the filling of the entire n layer with electron-hole plasma, a current, determined by the external circuit, with density j. goes through the

structure. At the initial stage of the plasma dispersion, the current of the nonequilibrium carriers by the p^+-- n junction is a diffusion current. The characteristic size of this region $L_d \approx qD_p p_m/j \leq 10^{-4}$ cm for typical values $p_m < 10^{16}$ cm^{-3} and $j > 10^2$ A/cm^2. After time $\tau_d = qp_mL_d/j = 10^{-9}$ s, the concentration of the electrons and holes at the p^+-- n junction drops to a value at which the transfer will have a drift character, and the electrons will begin to depart from the p^+-- n junction to the depth of the n layer with a sped

$$v_{nn} = \frac{1}{q(p_m+n_m)}. \quad (5.31)$$

In the diode structure by the p^+-- n junction, there arises an SCR (expanding at the very same rate, that is, $a = v_{nn}t$) at which the voltage will be restored:

$$U_{SCR} \approx \frac{qp_0a^2}{2p}. \quad (5.32)$$

where $\rho_0 = N_d + j/(qv_{sp})$, and v_{sp} is the saturated drift speed of the holes.

In a transistor structure, the electrons which left the n^{++} layer earlier, flying through the part of the p^+ layer which remained free of them up to the moment of switching during time

$$\tau_{np} = \frac{(w_p-L_{pK})}{j}qN_d, \quad (5.33)$$

reach the p^+-- n junction, and, compensating for the space charge with density ρ_0 in the SCR, they curtail the increase in the voltage on it.

From expressions (5.31) - (5.33), we find that, by the moment of the end of the flight electrons across p^+ base

$$U_{SCR} \approx U_m\frac{N_d}{N_a}. \quad (5.34)$$

The obtained evaluations show that in the transistor structure, the increase in the voltage on the SCR after the run of the ionization wave can be slowed down significantly. The sign of the equality in expression (5.31) corresponds to the optimum case. After all of the holes are recombined in the n layer or (and) leave the n layer for the p^+ layer and recombine there, the current through the structure and the arrival of the electrons from the n^+ layer are stopped, and at the p^+-- n junction, restoration of the voltage begins

The information given above was proven on n^{++}-- p^+-- n -- n^+ structures, $N_d = 10^{14}$ cm^{-3}, $w_p = 15$ μm, $w_n = 250$ μm, satisfying equation (5.31). The experiment have shown, that in a transistor structure, the stage for the restoration voltage is absent, as distinguished from the diode; moreover, after switching, the voltage smoothly drops during the entire time that the current flows, which is 10 ns, and is determined by the pulse generator used.

5.5 Thyristor like switches

In a number of cases, it is necessary that, after fast switching, the switch remain in the conducting state for an unlimited length of time until the current through it is not broken at the necessary moment by the compulsory method. In order to realize this switch in the transistor peaker discussed above, it is necessary to eliminate the possibility of carrying out the accumulated holes; that is, to guarantee the complete regeneration of the hole current. It is simpler than any way to do this after adding an additional p^+ emitter to the transistor structure from the direction of the n layer; that is, creating a thyristor structure. However, the switching characteristics of such a structure turn out to be unsatisfactory for the following basic reasons. When there is a rate of $U' \approx 10^{12}$ V/s, the complete time for raising the voltage to 3 kV equals approximately 3 ns, while the density of the displacement current provides in the NR of the n base a field intensity at which the drift sped of the carriers is close to the saturated speed (10^7 cm/s). Therefore, the holes injected from the p^+ layer to the n layer pass an n layer having a thickness of 200 μm, and they end up in the SCR earlier than the voltage will be created in it that is, they do not make it possible to delay ionization. In addition, due to the high

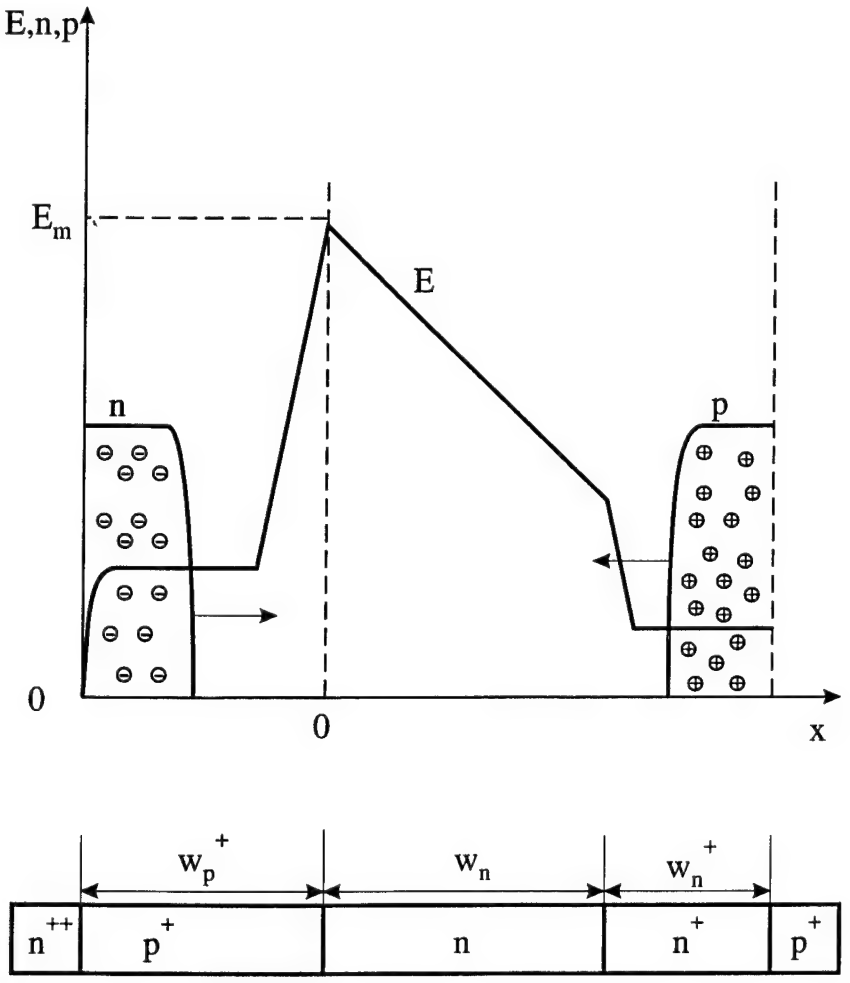


Fig.5.12
Thyristor with delayed ionization.

gain coefficient of the p -- n -- p transistor, large collector leakage currents arise, which, as was shown previously, also prevent the delaying of ionization.

It is possible to avoid these shortcomings in n^{++}-- p^{+}-- n -- n^{+}-- p^{++} type structures (fig. 5.12), in which, between the p^{+} emitter and the n base there is introduced an n^{+} layer which provides the additional delay of the injection of holes to the n-base. The ratio of the parameters of the p^{+}-- n section are the very same as in the transistor structure described above [see expression (5.31)]. The parameters of the n^{+} layer should guarantee a delay of the flight of the holes τ_p which is equal to the delay of the electrons in the p^{+} layer:

$$\tau_p = \frac{\mu_n q w_{n^+} N_{n^+}}{\mu_{pj}}, \tag{5.35}$$

where N_{n+} is the concentration of the donors in the n^{+} layer.

From here, taking expression (5.33) into consideration, we get

$$w_n N_{n^+} = N_a(w_p - L_{nc})\mu_p/\mu_n \ . \tag{5.36}$$

That is, the parameters of the p^{+} layer are close to the parameters of the n^{+} layer.

The experimentally obtained curves for the change in the voltage and current in such a structure almost repeat those for the transistor structure. However, it is possible to bring the n^{++}-- p^{+}-- n -- n^{+}-- p^{++} structure out of the switched on state only by means of decreasing the current through it to a value lower than the confinement current.

The turn on times of both transistor and thyristor are the same as of SAS (<0.3ns).

5.6 The possibilities of the numerical modeling of delayed impact ionization

In chapter 5 a new phenomenon is described -- that of super-fast switching in p -- n junctions. The most unexpected property consists of the fact that super-fast switching can occur when there is a relatively slow increase in the voltage, during a time which is a lot greater than the time of the voltage drop during switching . In this case, the current density increases homogeneously enough along the entire area of the structure. In order to explain this characteristic, a two-stage switching mechanism is assumed. At the initial stage in the raising of the voltage, unbalanced carriers, which are holes, arise due to impact ionization by the main carriers in the NR. These nonequilibrium carriers are carried out into the expanding SCR, and, falling into the super-voltage region where the field intensity is many times greater than the threshold field intensity, they excite a fast ionization wave. The calculations done for such a two-stage mechanism turned out to be in good qualitative and satisfactory quantitative agreement with the experiment when there were large SCR dimensions (more than 10^{-2} cm) and average U' values (10^{12} V/s). In the area of large U' values which are close to 10^{13} V/s and small SCR dimensions (less than 10^{-2} cm), the switching process can go in another, one-stage, route, when the formation of a overvoltaged region occurs on the background of a relatively large flow of initiating carriers. Naturally, the gain with respect to fast response in a one-stage process turns out to be less. The minimum switching time can be less than 0.05 ns.

We should pause in somewhat more detail on the problem of the accuracy of the calculation of the possible models of the switching process. The description, which is the most complicated for analysis, of the ionization wave itself, which is fundamentally nonlinear, even within the framework of the hydrodynamic approach requires the solution of a system from the Poisson equation (taking into consideration the charge of the mobile carriers) and the continuity equations (taking into consideration the dependence of mobility on the field intensity). A significant feature of the problem is the fundamental necessity of giving an integral (with respect to the entire thickness) condition for the distribution of the field (the Kirchgoff equation), since the influence of the external circuit. Therefore, during the analytical solving of the problem, it is impossible to use the methods which are well developed for nonlinear wave processes, also including the automodel solutions which are widely used during the calculation of a TRAPATT system of avalanche flight diodes. An attempts were made

to study the process of switching a p -- n junction by means of computer modeling in a one-dimensional approximation by Megapulse team and by others]. In all cases, the calculations led to the following, generally speaking, obvious results. When using only sufficiently accurately established coefficients for zone -- zone impact ionization for obtaining a noticeable gain with respect to fast response, a small leakage current was needed which provided an initial concentration of less than 10^6 cm^{-3}. Then the average distance between the carriers was more than 10^{-2} cm, that is, it was comparable to the characteristic dimensions of the structure. It is obvious that the hydrodynamic approximation which was accepted in the modeling for these conditions is already unusable, and a discretion consideration is necessary.

In addition, in order to obtain the small switching time observed in the experiment (0.2 ns and less), the calculated voltage on the structure, and consequently the field intensity on it as well, should be significantly (1.5 to 2 times) greater than the experimental values.

From what has been previously stated, it follows that both impact and tunnel ionization through deep levels can play a significant role in the switching process. However, up to the present time, these processes have practically not been studied, and can not be put in the form of concrete formulas with numerical coefficients in the model. In addition, in the case of especially fast models, an increase in the spatial instability (nonhomogenity) is possible. In this case, the one-dimensional approach is generally not usable. In those cases where, for the numerical calculation, there are sufficiently reliable data (data on the initial stage up to the formation of the wave), the simple analytical approach developed in section 5.2 gives a good coincidence with computer modeling.

In this chapter, the problem of the resolution of plasma after switching has almost not been discussed at all, since in diode structures this process is similar to the processes previously described in chapter 4.

6 Investigation of properties and limitations of pulse forming semiconductor networks with its application to more than 100KV design.

6.1 Peak power and fronts

6.1.1 General consideration

All methods of power short pulse generation by use of the new switches are based on time compression of initially long (submicrosecond) pulses. Generally speaking, output pulse power is determined by the product of power of single switching devices and the number of the devices connected in series and/or parallel in power compressing network. In this case two main questions arise: what is the maximum possible power of the switching device, how to sum powers of many devices and what is the most effective means to sum powers?

It is clear that the answers to the questions depend on the types of devices and are different for drift step recovery devices or silicon avalanche shapers. Nevertheless some remarks common for both types should be made. Almost all new devices (excluding DSR transistors and thyristors) are two terminal devices. This is the feature, that determines their use in power compressing networks. It is evident, that for the most simple cases, when only one switching device is used (see fig. 1.5 and fig. 1.9), such circuits are more complicated than well known circuits using three-electrodes switch triggered through the third electrode. In the last case pulse forming network has far less number of additional components (parts).

When a large number of switching devices must be used for power increase, the case of two electrode devices becomes more simple, because the devices require no triggering circuits. For example, many DSRD wafers may be

soldered one on the top of other (stacked). Such stack may have ten or more times higher voltage, than single diode, but for an end user the stack appears as one device or single unit having bigger thickness.

Such possibility of combining of both types devices DSRD and SAS has been discussed in parts 4.3.3 and 5.3.2.

Here, we will consider the limitations of the combining of a large number of devices in a "stacked" manner.

6.1.2 Drift step recovery devices

It was shown (see formula (4.4), that maximum rate of voltage increase limitations for DSRD is of fundamental nature and is near $U \approx 2 \cdot 10^{12}$ V/s. From the limitation it follows: the shortest pulse front, generated by DSRD (τ_f) is

$$\tau_f \geq \frac{U_d}{U'} = \frac{E_\alpha \cdot W_n}{2U'} \qquad (6.1)$$

where $U_d = E_\alpha \cdot W_n/2$ is the maximum diode blocking voltage, E_α - threshold ionization field, W_n - n-layer width (the best efficiency case is $W_n = \frac{\varepsilon E_\alpha}{qN_d}$,when all n-layer is overlapped by space charge region in off-state).

It may be noted that for quasisimmetrical diode U' may be two times more (up to $4 \cdot 10^{12}$ V/s).

Equation (6.1) shows that 2 kV diode can generate 1 ns front, 2 kV amplitude pulse.

For the best efficiency case it follows from (6.1)

$$\tau_f = \frac{\varepsilon E_\alpha^2}{2qN_d} = \frac{\varepsilon E_\alpha^2 V_s}{2j_-} \qquad (6.2)$$

The condition of the fulfilling of (6.1) $j_- = qN_dV_s$ was taken into an account.

We should like to emphasize that the switching off process may be slowed down in comparison with (6.1) (as was shown in the part 4.2.3) by simple increase of the diode area. In this case current density will be less than saturated one: $j_- << qN_d V_s$. Two general important rules follow from (6.1) and (6.2) .

1. The less front of generated pulses with the same amplitude U_0 , the larger number of wafers (N) can be used in the stack

$$N \approx \frac{U_0}{U_d} = \frac{2U_0}{E_\alpha \cdot W_n}, \qquad (6.3)$$

2. The less front, the higher current density can be used.

As was mentioned before, the good uniformity (synchronism) of voltage recovery on each diodes can be reached in case of good reproducibility of diodes parameters (N_d, S). From (4.3) it may be shown,

$$\delta U \approx \delta N_d$$
$$\delta U \approx \delta S, \qquad (6.4)$$

where δU - relative dispersion of the diode voltage at the given moment, δN_d, δS - relative dispersion of doping concentrations and areas. One may conclude from (6.4) that good control of diode parameters (the requirement of large life-time of minority carriers must be added as well) permits "infinite" increase of diodes number in the stack. Nevertheless the increase of the diode stack voltage increases the thickness (height) of the stack. From (6.1), (6.3) it follows that the minimal thickness of the stack (W_s) of N diodes

$$W_s \geq \frac{2U_0}{E_\alpha} \qquad (6.6)$$

is independent of front duration.

In the case of short fronts, the stack consists of the larger number of the thinner diodes. Actually the thickness of heavy doped layers, contacts and soldering interface must be taken into consideration. Due to this addition, actual thickness may be times more than minimal (6.6) and short front stack should be considerably thicker than the long front one.

The thickness increase causes the increase of inductance of the stack that may worsen the front. For example, the thickness of the 100 KV stack estimated by (6.6) is more than 1 cm. Taking into an account the mentioned above additions the thickness can reach 4-6 cm.

As was mentioned in part 2.3 (see fig. 2.2), the switch should be connected in a circuit as a part of a waveguide. For such a connection the next limit for a stack diode thickness (W_s) may be evaluated:

$$\tau_f >> \frac{L}{\rho} \cong \frac{W_s}{c_w} \tag{6.7}$$

where c_w - electromagnetic wave velocity in the waveguide, L - inductance of the stack, ρ - waveguide impedance.

In spite of short duration of high voltage pulses, oil or other dielectric filling should be used in high voltage pulsers. In the case of oil filling ($c \approx 2 \cdot 10^{10}$ cm/s), we get from (6.6), (6.7) the upper voltage limit for a stack with 1 ns turn off time.

$$U_0 \lesssim \frac{W_s E_\alpha}{2} \approx \frac{E_\alpha \cdot c_w \cdot \tau_f}{2} = \frac{c_w E_\alpha \cdot W_n}{2V_s} \tag{6.8}$$

$$U_0 \lesssim 1 \text{ MV}.$$

Let us consider the upper limit of the braked current (switched off). As was shown before (2.34) for injection or extraction mechanism of switching the maximum area of a semiconductor wafer of the round shape (S) limited due to skinning is

$$S \lesssim \frac{c^2 \cdot W^2}{V_s^2} = c^2 \cdot \tau_f^2, \tag{6.9}$$

where c is the velocity of electromagnetic wave in semiconductor.

The maximum current is proportional to the front duration:

$$I_m \lesssim j_s \cdot S = c^2 \varepsilon E_\alpha \tau_f, \tag{6.10}$$

The stability of the uniform current density distribution across the device area was discussed in parts 4.3 (for DSRD) and 5.1 (for SAS). The stability conditions for SAS were obtained and was shown, as well, that uniform stable current distribution in DSRD is on inherent peculiarity.

The maximum power commutated by single stack of wafers (P_m) may be deduced from (6.8), (6.10)

$$P_m \lesssim c_w c^2 \varepsilon E_\alpha^2 \tau_f^2. \tag{6.11}$$

For $\tau_f \approx 1$ ns (6.10) and (6.11) yield: $I_m < 10^4$ A, $P_m < 10^{10}$ W.

It should be noted that the actual limitation on the switching power of one stack of wafers strongly depends on the device technology, additional requirements on heat dissipation and so on.

6.1.3 Silicon avalanche shapers

SAS is a two-electrodes device as well and everything written above in part 6.1.1 about advantages and disadvantages of two-electrodes switches are valid for SAS. For output power increase many SAS wafers should be assembled in a stack as well. The same as for DSRD limitation (6.7), (6.8) on the thickness of the stack is valid, but the parameters E_α and τ_f for SAS are different.

As was shown in part 5.3.2 current level of understanding of delayed ionization switching is not so good, as for fast recovery processes and it is possible to make only rough estimations for τ_i.

The maximum field intensity and voltage for SAS are higher than for DSRD. The upper limit for E_α is $E_\alpha \lesssim b$, where b - is the threshold field in (5.3). As was shown in part 5.3.2 the actual value of E_α is saturated in Si at $\sim (4 \div 5) \cdot 10^5$ V/cm level, two-three times less than b, but still two-three times more than static breakdown threshold. So, the difference for E_α in cases of SAS and DSRD is exists, although it is modest.

It is evident, that turn on time of SAS (τ_f) is not less than the product of the mean time of one act of the impact ionization (τ_i) and of the multiplier determined by ratio of concentration of carriers at the initial moment (n_i) and the final concentration (n_m)

$$\tau_f \gtrsim \tau_i \ln \frac{n_m}{n_i} = \frac{1}{\alpha_{(E\alpha)} V_s} \ln \frac{n_m}{n_i}. \qquad (6.12)$$

From (6.12) turn on times $\tau_f \approx (10 \div 100)$ ps can be evaluated. It should be reminded that the turn on time is determined by the dU/dt applied to the device, as was shown in part 5.3.2.

From (6.8) we get:

$$U_{0m} \lesssim 200 \text{ kV for } \tau_f \approx 100ps$$
$$U_{0m} \lesssim 20 \text{ kV for } \tau_f \approx 10ps.$$

As was mentioned in part 5.1.1, the delayed ionization can be subjected to instabilities, determined by the fact that the increasing of the field intensity decreases the time of ionization and increases the rate of carrier generation. When many devices are assembled in a stack the same fact helps to synchronize the processes of switching of the devices. Let us suppose that due to some initial deviation of a diode parameter (for example the smaller diode thickness) this diode switches on faster than others. Faster voltage drop on the device increases the voltage drop on the other devices. The increasing of the voltage accelerates turn on processes and "slow" devices catch up the fast one. So the requirements for the uniformity of the assembled in a stack SASs parameters are weaker even than in the case of DSRDs.

As was shown in the part 2.3 (2.35) the skin effect limits the area of semiconductor wafer to the value

$$S < \frac{c^2}{V_s^2 \alpha_{(E\alpha)}^2} = c^2 \tau_i^2 \ . \qquad (6.13)$$

Formula (6.13) in appearance is similar to (6.9). The maximum current density in turn on state (j_m) is determined by the maximum concentration in accordance with (5.3)

$$j_m = q V_s n_m = \frac{\varepsilon E_\alpha}{\tau_i} \left(\frac{E_\alpha}{b} \right) . \qquad (6.14)$$

From (6.14) we get

$$I_m \lesssim c^2 \varepsilon E_\alpha \tau_i \left(\frac{E_\alpha}{b} \right). \qquad (6.15)$$

Again the limit (6.15) is very similar to the limit in the case of DSRD (6.10). From (6.15) we get:

$$I_m \lesssim 2 \cdot 10^3 A \text{ for } \tau_f \approx 100 \text{ ps, S} \approx 1 \text{ cm}^2$$
$$I_m \lesssim 200 \text{ A for } \tau_f \approx 10 \text{ ps, S} \approx 0{,}01 \text{ cm}^2$$

Finally we evaluate for the peak power ($P_m \approx U_{0m} \cdot I_m$) : :

$$P_m \lesssim 4 \cdot 10^8 \ W \text{ for } \tau_f = 100 \text{ ps.}$$
$$P_m \lesssim 4 \cdot 10^6 \text{ W for } \tau_f \approx 10 \text{ ps.}$$

The evaluated power of a single diode stack is not the limit for a pulser output.

The next step in power increase (higher than the limits shown) is to sum powers of many such stacks into a single load.

The limit to the stack voltage implies that in the thicker stack the diodes, placed on the opposed ends work causally independently. The delay of the electromagnetic wave propagation along the stack is more than the duration of the switching processes.

It may be shown, that the same causal independence of the processes at the opposed edges of the wafer exists when the device area is larger than limited by (6.9), (6.13).

Therefore the summing of powers of the stacks is possible only in the case of external synchronization of the pulses generated by each stack.

The possibility of synchronization a large number of independent sources and the stability of delays of power compressing cells will be considered later. Then the means of power summing up by use of transmission lines will be considered as well.

In conclusion, brief mention about the pulse width (FWHM) should be made. For given switching time of the device, the width of the generated pulse, at the most degree is determined by the circuits (pulse forming network - PFN). Nevertheless the requirement of the best efficiency imposes sever limitations to the shortest and the longest possible pulses.

For the case of short pulses the limit is very simple and clear: turn on (or off) time is equal to the pulse front, is equal to the pulse decay and is equal to FWHM. When FWHM is shorter than turn on or turn off time the efficiency drops drastically.

For the case of the long pulses the limits are determined by either the device's parameters or the actual network used. The example of a network for generation of step pulse with practically infinite width was shown in part 1.2.1 (thyristor-diode closing switch). Some examples of the pulses of different shapes will be considered later.

6.2 Electrical efficiency

6.2.1 General consideration

Total electrical efficiency of pulsers determined as a ratio of output and input energy, strongly depends on the form of the pulse and the type of the circuits. For example, it is evident that in the case of a very long step-like pulse the commutation losses play no role.

For very wide range of use, the main requirement is to put as much energy as possible into the shortest time period. It is evident (and was mentioned before) that the good efficiency is possible only when the switching time of the switch is shorter than FWHM. The mentioned conditions determine the shape of the pulse - it is bell like pulse. The most efficient way to generate such pulses by use of two terminal switches is the pulse compressing networks (the examples of such networks have been considered in part 1.2). The pulse compressing network could include many stages consisting of simple compressing cell.

The total efficiency of the network (ξ) is the product of partial efficiencies (ξ_i) of each stage of power compression (cell) when several stages are used.

$$\xi = \xi_1 \cdot \xi_2 \cdot \ldots \cdot \xi_i .$$

Each cell may be characterized by the power compression ratio (η_i) which is the ratio of the short pulse power (the output of the cell) and the initial long pulse generated inside the cell by relatively slow switch (see fig. 1.5).

The semiconductor compressing cells, using DSRD and SAS, have limitations to the maximum compression ratio, considered in the parts 4.3, 5.3. The actual number of compression cells is determined by such requirements to the pulser under design as following: efficiency, power, size, weight, cost and so on. At the current time no regular procedure exists to get an optimal number of cells and compression ratio distribution between the cells.

Our experience shows that for the purpose mentioned above two stages using DSRD and one or two stages using SAS may be used.

Let us consider the network consisting of two DSRDs stages and one SASs, shown in fig. 6.1. The energy losses for one cycle of compression may be separated into two types:

a) energy dissipation into the heat in each part of the circuit during the time of the conductivity current flow

b) even in the case of no conductivity current only part of energy stored initially in the cell can be compressed and thrown out into the next stage or the load. Generally speaking the remained part of energy could be recaptured (recuperated) and used in the next cycle of pulse forming, but practically in nano and subnanosecond time band the current level of technology

many stages power compressing network

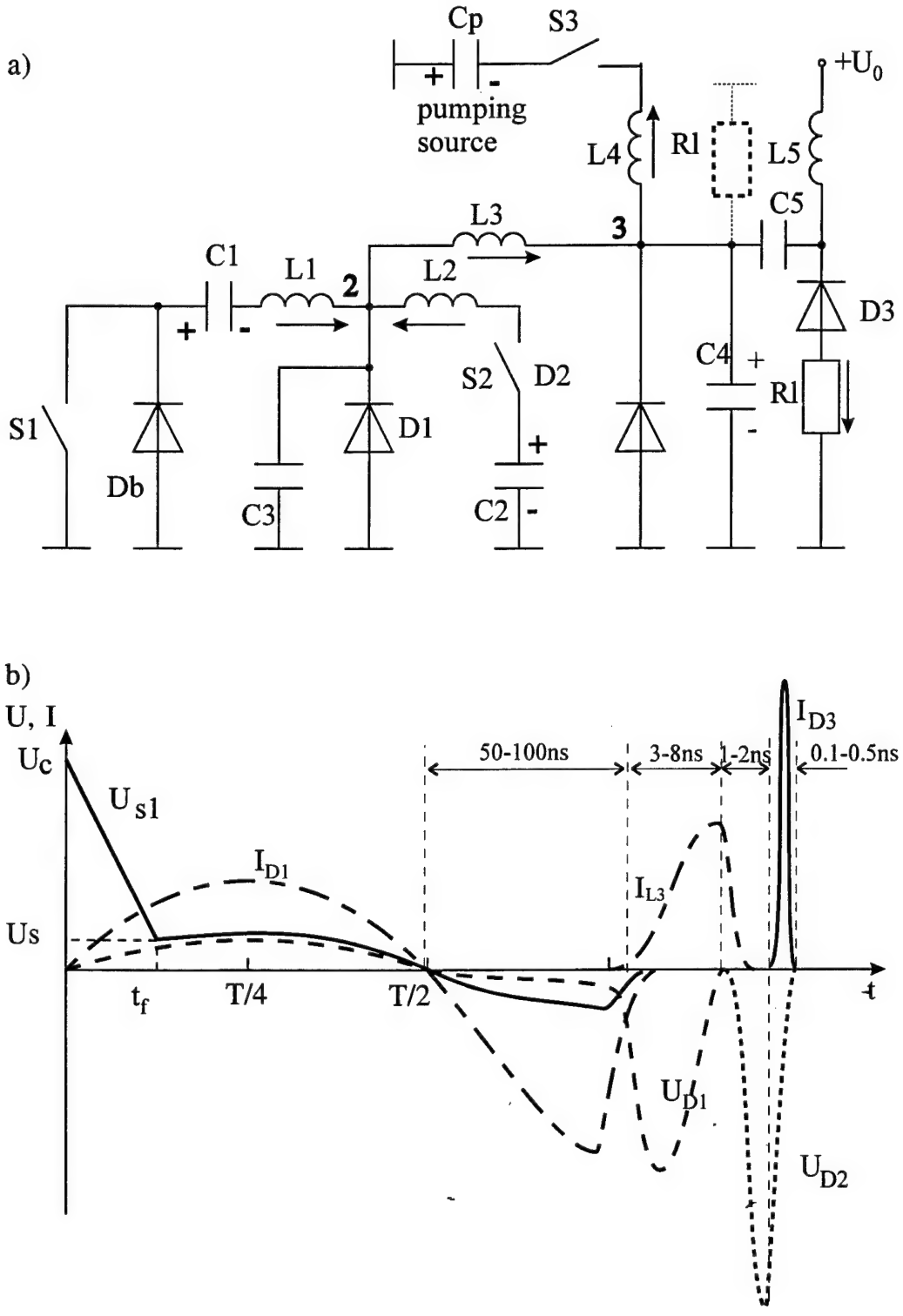


Fig.6.1

does not permit such recuperation. Finally, the remained part of energy either dissipates into the heat after several cycles of oscillations or produces a (parasite) after pulse some time later after the useful one.

The first stage of compression (S_1, C_1, L_1, L_2, C_2, S_2, D_1) is similar to considered in the part 1.2 (see fig. 1.5). The second DSRD (D_2) is pumped by additional pulse forming circuits C_p S_3 through the inductor L_4. The output SAS (D_3) is biased by DC voltage from the source U_0 through the inductor L_5. When the first DSRD (D_1) turns off, the energy stored in the inductors L_1, L_2 is transferred to charge capacitance of the diode D_1 ,the additional capacitor C_3 and to increase the current of the second stage storage inductor L_3. It should be noted, that due to the initial pumping, the second DSRD (D_2) still is in conducting state in spite of the reverse current.

It will be shown later, that in the case of high Q factor of the first stage and under conditions $L_1 = L_2$, $C_1 = C_2$, the most of the energy stored in L_1, L_2 will be transferred to the storage inductor L_3, after discharge of C_3 if the condition $L_3 = L_1/2$ is fulfilled. The capacitor C_3 (the capacitance of D_1 must be added as well) determines only the delay of energy transfer and the maximum voltage at D_1.

When the current in the inductor L_3 reaches the maximum value, the diode D_2 turns off (the charge stored in D_2 should be adjusted by pumping source to guarantee the mentioned condition). The energy stored in L_3 is transferred to the diode D_2 capacitance, the additional peaking capacitor C_4 and capacitance of SAS (D_3).

When the diode D_2 voltage reaches the threshold voltage of SAS (D_3) switching on, the capacitor C_4 and the capacitance of DSRD are discharged into the load Rl.

The typical times (half period τ) of the fig. 6.1 network are:

1. circuit L_1C_1 L_2C_2 $T=\tau = \pi\sqrt{L_1 \cdot C_1} \approx 0,1 \div 0,5\ \mu s$
2. circuit L_1C_3; L_2C_3 $\tau_f = \tau \approx \pi\sqrt{L_1 \cdot C_3} \approx 5 \div 20\ ns$
3. circuit L_3C_4 $\tau_f = \tau \approx \pi\sqrt{L_3 \cdot C_4} \approx 0,5 \div 2\ ns$
4. circuit C_4R_l $\tau \approx C_4 R_l \approx 0,1 \div 1$ ns.

The energy passing through each stage can not be less than the output energy in the load. It may be shown (see (6.11)) that the output energy (Q) of the cell which compresses pulse down to 1 ns FWHM by use of single DSRD stack is limited by the value $Q \approx 10$ J.

For the pulse having 0,1 ns FWHM we get from (6.8) and (6.15) $Q \approx 4 \cdot 10^{-2}$ J. This levels of energies should be taken into an account for further detailed consideration.

6.2.2 Submicrosecond circuits

Initially total energy is stored in the capacitors C_1 and C_2 charged up to the voltage U_C. When the switch S_1 closes, the oscillations in the $S_1L_1C_1D_1$ circuit start. During the first halfperiod the current flows through the diode D_1 in the forward direction. The most energy losses are connected with semiconductor devices: closing S_1 and opening D_1 switches. Quality factor Q_L of the value more than 50 for ironless inductors L_1, L_2 may be easily reached. Q-factor for small ceramic capacitors may be as high as 30÷50. During 3/4 of LC cycle (total operational time of the cell), total energy losses in L and C are less than 3% and will not be taken into account further. The losses connected with the closing switches may be separated into two parts - commutation losses (Q_C) during turn-on time and residual losses (Q_S) determined by residual voltage in "on" state. Transient switching process is illustrated at fig., 6.1 (b): during turn on time (τ_{fs}) voltage drop on the switch decreases from the initial stationary value U_C down to the sustaining value U_S. The last may be characterized by on state resistance R_C.

To estimate commutation losses we consider simplified case when the voltage decrease is linear. Computer modeling has shown that the difference between linear type of the voltage decrease and nonlinear one is minor (less than two times).

In LC circuit the current shape is sine-like. For this case we get

$$Q_C = \int_0^{\tau_f} \frac{U_0}{\tau_{fs}}(\tau_{fs} - t) I_m \sin\varpi t \, dt, \tag{6.16}$$

where ϖ - frequency of L_1C_1 circuit, I_m - maximum pumping current.

When turn on time τ_{fs} is far less than the halfperiod ($\tau_{fs} << \frac{1}{\varpi}$), (6.16) may be simplified

$$Q_C \cong \frac{U_0 I_m^2 \varpi \tau_{fs}^2}{6} \tag{6.17}$$

For the residual losses we can write

$$Q_S = \int_0^{\pi} \frac{R_S \cdot I_m}{\varpi} \sin^2 \varpi t d\varpi t = \frac{\pi R_S I_m^2}{2\varpi} \tag{6.18}$$

Taking into an account that $I_m = \frac{U_C}{\rho}$ (where $\rho = \sqrt{\frac{L_1}{C_1}}$ is L_1C_1 impedance), we get from (6.17) (6.18).

$$Q_C \approx \frac{Q_0}{3}(\varpi\tau_{fs})^2, \; Q_S = \frac{Q_0 \cdot R_S \cdot \pi}{\rho}, \tag{6.19}$$

where $Q_0 = \frac{C_1 U_C^2}{2}$ is the energy stored in C_1.

From (6.19) it follows, that the commutation losses depend very strongly (square-like) on the turn on time and the residual loses are proportional to "on" resistance. During the second halfperiod the current passes through the additional "bypass" diode D_b (see fig. 6.2). It may be shown, that losses at D_b are small part of the losses on DSRD.

In part 4.3.2 it was shown that for DSRD the condition $\frac{\sqrt{DT}}{W} \lesssim 1/5$ (where $T = \pi/\varpi$ is halfperiod) should be fulfilled. From this condition we can get

$$T \lesssim \frac{W^2}{25D} \approx \frac{(\tau_f V_s)^2}{25D} \tag{6.20}$$

The condition (6.20) shows that the faster DSRD, the less permitted halfperiod. As it follows from (6.19), for the given turn on time of the primary switch (τ_{fs}) and turn off time of DSRD (τ_f) the decrease of T (increase of ϖ) improves the switching properties of DSRD, but worsens the efficiency of the primary closing switch. Two stage network considered here permits to use slow DSRD for the first stage with good efficiency to get short halfperiod for the second stage with fast DSRD.

Hence the second $L_2C_2S_2$ circuit is switched on after halfperiod, the current flows in the circuit only during a quarter of a period. The commutation losses remain the same as in the first L_1C_1 circuit, but the residual loss is two times less than in (6.18) and for total losses in closing switch (Q_{ct}) we can get

$$Q_{ct} = Q_{0t}\left(\frac{(\varpi\tau_f)^2}{3} + \frac{0{,}75 \cdot R_S \cdot \pi}{\rho}\right), \tag{6.21}$$

where $Q_{0t} = 2Q_0$ - total energy initially stored in both LC circuits.

After 3/4 of the period of LC circuits total energy (minus losses) is stored in the inductors L_1 and L_2. If the losses in the switches S_1, S_2 are small ($Q_{ct} << Q_{ot}$), the currents in L_1 and L_2 are equal during the second halfperiod. The diode D_1 current is the sum of L_1 and L_2 currents and the diode turns off when the total current reach the maximum value. As was mentioned before, when D_1 is turning off the L_1 and L_2 currents are charging the capacitor C_3 and the capacitance of DSRD. Simultaneously these currents increase the current in the inductor L_3.

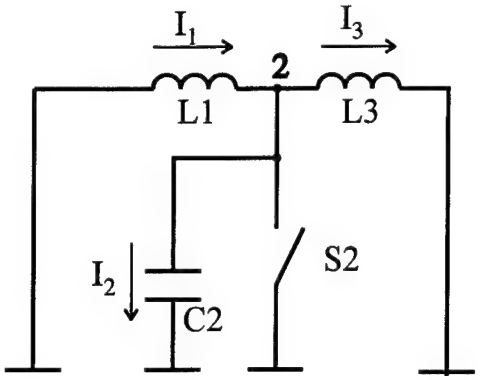


Fig.6.2

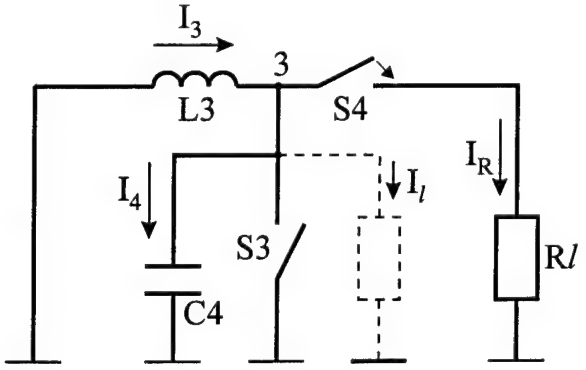


Fig.6.3

6.2.3 Nanoseconds cells

The process of the energy (current) transfer from the inductors L_1 and L_2 to the inductor L_3 is presented in simplified version in fig. 6.2. Two inductors L_1 and L_2 from fig. 6.1 are represented as one storage inductor L_1 with the current I_1, which is equal to the sum of the currents in L_1 and L_2. The diode D_1 is modeled by the opening switch S_2 and the capacitor C_3, which (in linear approximation) includes the capacitance of D_1 and possible additional external capacitor C_3 (shown in fig. 6.1). In the part 4.2.3 it was shown (see (4.39)) that the voltage restoration on DSRD may be represented as the charging of a nonlinear capacitance of space charge region and of a geometrical capacitance of the diode. Computer modeling shows, that the linear approximation considered here conserves the main features of the processes under investigation. It may be shown that the made representation (4.39) does not contradict the representation of switching as a widening of SCR used in (6.1). Both representations are different views of the same physical phenomenon. Initial conditions for circuit on fig. 6.2 are

$$t = 0, \quad I_1 = I_m, \quad I_3 = 0, \quad U_2 = 0$$

Where U_2 is the capacitor C_3 voltage (node 2). It may be shown that the currents and voltage in the circuit are described by the next set of equations:

$$U_2 = \frac{I_m \cdot \rho_n \sin \varpi_2 t}{\sqrt{\left(1+\frac{L_1}{L_3}\right)}} = \frac{I_m}{C_3 \varpi_2} \sin \varpi_2 t \qquad (6.22)$$

$$I_3 = \frac{I_m(1-\cos \varpi_2 t)}{\left(1+\frac{L_3}{L_1}\right)}$$

$$I_1 = I_m\left(1 - \frac{(1-\cos \varpi_2 t)}{\left(1+\frac{L_3}{L_1}\right)}\right)$$

where $\rho_n = \sqrt{\frac{L_1}{C_3}}$, $\varpi_2^2 = \frac{1+\frac{L_1}{L_3}}{L_1 C_3}$.

Thus we have that if $L_1 = L_2$ at the time $\tau_t = \pi/\varpi_2$, $I_1 = 0$, $U_2 = 0$, $I_3 = I_m$, total energy stored in the inductor L_1 is transferred to the inductor L_3 .

In the case of $L_3 << L_1$ the current transferred into L_3 may be two times more than the initial current I_1. So, not only voltage, but the current "multiplication" is possible in compressing network with DSRD. The pay for the multiplication is the bad efficiency which in accordance with (6.22) drops to zero level for the current dubbing ($I_1 = -I_m$, when $L_3 = 0$, $\pi = \omega_2 t$).

In the case of the poor Q-factor, due to losses in S and D_1, the current's symmetry of the first cell (fig.6.1 ; C_1 L_1 C_2 L_2) is broken. The currents(at the moment of their maximum value) of the inductors L_1 (I_{1m}), L_2(I_{2m}) are not equal ($I_{1m} < I_{2m}$). A part of the current I_{2m} is transferred into the inductor L1 when DSRD brakes the current. It may be shown, that when the current of the inductor L_3 reaches the maximum value, the currents in L_1 and L_2 are:

$$I_{10} = (I_{1m} - I_{2m})/2 = I_{1m} - 2\square(I_{1m} + I_{2m})/L_1\square$$

$$I_{12} = (I_{2m} - I_{1m})/2 = I_{2m} - 2\square(I_{1m} + I_{2m})/L_2\square$$

These expressions confirm the conclusion made in the part 1.2.1: to decrease the residual currents it is necessary to adjust the ratio of the inductors $L_1 > L_2$, when $I_{1m} < I_{2m}$.

The energy connected to this currents (Q_b) is lost

$$Q_b = \frac{(I_{1m} - I_{2m})^2}{(I_{1m}^2 + I_{2m}^2)} Q_0 \ , \qquad (6.23)$$

where Q_0 is the energy initially stored in L_1 and L_2.

In accordance with (6.22) the maximum voltage at the opening switch U_2 decreases when the period τ_t of energy transferring increases.

At the preceding part we did not considered the losses in DSRD during the first submicrosecond cycle of power transferring from capacitors C_1 and C_2 to the inductors L_1 and L_2 (see fig. 6.1). Now we see that during the next nanosecond cycle of power transferring to the inductor L_3, the voltage on DSRD (D_1) in off state reaches the high value ($U_2 >> U_C$).

Therefore the residual voltage on the DSRD's stack in conducting state may be high enough and the energy losses in the stack may play a significant role. The voltage drops on a diode for pumping and extracting phases were determined in (4.21) and (4.31). The equations (4.21) and (4.31) have been found for the case of the step like current. In our case of the sine like current they may be used as the upper bound. A number of diodes in stack (N) is determined by maximum value of U_2 from (6.22) and is similar to (6.3).

The requirement of the shortest turn off time, used in the part 6.1, now may be omitted and the current density far lower than saturated j_s may be used ,which helps to decrease the losses.

The turn off time (τ_{off}) may be estimated from (4.2)

$$\tau_{off} \approx \frac{W_n N_d \cdot q}{j} = \frac{\varepsilon E_\alpha}{j}. \tag{6.24}$$

Using (4.21), (6.3), (6.24) we have for the voltage drop on the diode stack (U_s) in on state.

$$U_{s1} \lesssim \frac{4U_2 \cdot W_n}{3E_\alpha \cdot \sqrt{\tau_{off} \cdot T \cdot \mu_n \mu_p}}, \tag{6.25}$$

where $T = \frac{\pi}{\varpi}$ is halfperiod for the submicrosecond part of the compressing network (fig. 6.1).

It should be noted that voltage drop in diode during on state as may be seen from (4.21) and (4.31), insensitive to the current direction.

Formula (6.25) shows that voltage drop does not depend on the current. Actually, the current dependence is concealed in (6.24). For needed diode current I the area of the diode S is chosen to satisfy the expression (6.24) ($S = \frac{I}{j}$).

It should be noted that W and T are connected by condition (6.20) and can not be chosen separately. Substituting (6.20) into (6.25) we have

$$U_s \leq \frac{20U_2\sqrt{D}}{3E_\alpha \cdot \sqrt{\tau_{off} \cdot \mu_n \cdot \mu_p}} \tag{6.26}$$

The expression (6.26) shows that the turn off time (τ_{off}) should be as long as possible. The maximum value for τ_{off} is $\tau_{off} \approx \pi/\varpi_2 \approx \tau_t$. That means that the external capacitor C_2 is absent and only the intrinsic capacitance of the opening switch (DSRD) is used. It is well to bear in mind that although the last condition improves the efficiency in some cases a great deal of expensive semiconductor material may be needed. The estimation of energy losses (Q_s), derived from (6.26), is

$$Q_s \lesssim \frac{U_s I_m \cdot \pi}{(1+\frac{L_3}{L_1})\varpi_2} \tag{6.27}$$

The expressions (6.22), (6.26) show as well, that the increase of the halfperiod τ_t decreases maximum voltage U_2 and decreases losses. Nevertheless in the case of long τ_t two effects worsening the performance of the network start to play role (see fig. 6.1):

1. During the halfperiod τ_t the inductors currents recharge the capacitors C_1 and C_2. The charge returned to C_1 and C_2 is $\int_{\tau_t} Idt$ and may be derived from (6.22). The energy returned to the capacitors C_1 and C_2 (Q_r) is

$$Q_r = \frac{I_m^2 \tau_t^2 \cdot L_3}{(L_1+2L_3)(C_1+C_2)} = \frac{4Q_0 \cdot \tau_t^2 \varpi^2}{\left(2+\frac{L_1}{L_3}\right)}. \tag{6.28}$$

This energy can not be recuperated and will be lost.

2. The longer τ_t, the larger the charge that must be pumped into the second DSRD (D_2) (see fig. 6.1) and the pumping D_2 current. The impedance of the pumping source with large current capacity must be small. In this case noticeable part of the DSRD current may be diverted into the pumping source causing additional losses.

When the second DSRD D_2 breaks its current the current of L_3 charges the capacitance of DSRD D_2, peaking capacitor C_4 and the capacitance of SAS D_3.

If the pulser under consideration is designed to generate nanosecond pulse the load resistor (dashed lines) is connected in parallel with D_2 to the node 3. The other parts, shown on fig. 6.1 on the right-hand side of D_2, are excluded. In the last case part of the current of L_3 is diverted into the load resistor. When the voltage at DSRD reaches maximum value the capacitance of DSRD and of C_4 (if existing) are discharged into the load. Assuming the simplification made for fig. 6.2 this process of the pulse forming may be illustrated by fig. 6.3. The most interesting case for pulse forming is the case of unipolar pulse (no oscillation) having maximum amplitude on the load. It may be shown that for the case the next condition must be satisfied:

$$R_l = \frac{1}{2}\sqrt{\frac{L_3}{C_4}}, \tag{6.29}$$

and the pulse shape is bell like with FWHM $\approx 5\ R_l C_4$.

$$U_3 = \frac{I_m t \exp(-t/2R_l C_4)}{2C_4}, \tag{6.30}$$

where I_m - braked currant, C_4 - is the total capacitance of DSRD and of additional capacitor.

The pulse amplitude is

$$U_m \approx 0.7 I_m R_l \tag{6.31}$$

The energy accumulated in the diode capacitance C_4 is not lost and mostly is transferred later into the load.

As was shown in the part 4.2.3 the conductivity current exists only in the neutral region of DSRD during the phase of fast recovery. The losses Q_{nr} caused by the current are similar to the losses evaluated for SAS in the part 5.4:

$$Q_{nr} \lesssim \frac{E_s \cdot W_n \cdot 2U_m I_m}{V_s E\alpha} \approx \frac{E_s}{E_\alpha} Q_l, \tag{6.32}$$

where Q_l - energy, transferred to the load.

The condition of the shortest turn off time $j = j_s = qN_dV_s$ was used in (6.32). These losses are small: $\frac{Q_{nr}}{Q_l} \lesssim 5\%$. If the current is less than the saturated value j_s, the losses are smaller than in (6.32) proportionally j/j_s.

For estimation of the voltage drop on D_2 during "on" state, the expression (6.25) can be used with minor modifications

$$U_{S_2} \lesssim \frac{4U_m W_n}{3E_\alpha \sqrt{\tau_f \tau_l \cdot \mu_n \mu_p}}, \tag{6.33}$$

where U_m - output pulse amplitude, τ_f - front of the pulse.

When the D_2 capacitance and capacitor C_4 are used to store the energy the voltage at the D_2 (node 3) before the closing of SAS (D_3) is determined by

$$U_3 \approx I_3 \cdot \rho_3, \tag{6.34}$$

where $\rho_3 = \sqrt{\frac{L_3}{C_4}}$ and C_4 includes the capacitance of D_2 and D_3.

The losses in D_2 are determined by (6.32) (6.33), as at the previous case, but using the new value of U_3 from (6.34).

It may be shown that the energy, stored in nonlinear capacitance of space charge region (Q_{SCR}) is

$$Q_{SCR} = \frac{SqN_d \cdot W_{SCR} \cdot U_{SCR}}{3} = \frac{q^* \cdot U_{SCR}}{3} = \frac{C_4 U_d}{2}, \qquad (6.35)$$

where $C_4 = \frac{4\varepsilon S}{3W_S} = \frac{4\varepsilon S \cdot V_s}{3\tau_f}$ is the linear equivalent of the capacitance of SCR. W_S is the total thickness of all diodes in the stack (when W_n=W_{SCR}, i.e. SCR overlaps n-layer).

When this result is compared with that well known for linear capacitor, it is apparent, that the only difference is the multiplier 3/2.

6.2.4 Subnanosecond cells

When SAS (D_3 fig. 6.1) closes the capacitor C_4 discharges into the load. As was shown in the parts 5.3.2, 5.4 there is an optimal value for the voltage rise rate on one SAS ($dU/dt \approx 10^{12} \div 3 \cdot 10^{12}$ V/S depending on the type of SAS). The delay of the breakdown (turn on) is 1-2 ns. When many SASs are connected in a stack dU/dt value is increased proportionally to a number of the stacked SASs, but the delay keeps "one diode" short value.

The expression (6.32) determining the losses in DSRD during the phase of fast restoration is valid for the case of use of DSRD as a storage capacitors as well.

The maximum DSRD voltage (6.34) should match the maximum switching (turn on) voltage (U_{on}) of SAS. As practice shows to improve the switching properties (low resistance in turn on state, short turn on time , lower jitter at the moment of the switching on when there is non zero value of dU/dt), DSRD voltage should be slightly more than U_{on}. At the moment $t = \pi/3\varpi$ (where $\varpi^2 = 1/L_3 C_4$) the value of dU/dt is equal to 50% and the voltage is equal to > 85% of the maximum possible values (at the moments $t = \frac{\pi}{4\varpi}$ and $t = \frac{\pi}{2\varpi}$ consequently). If SAS voltage is matched to ensure switching on at the moment $t = \frac{\pi}{3\varpi}$, then only 75% of the energy stored in the inductor L3 is transferred to the capacitor C_4.

The energy remained in the inductor L_3 will not be lost, it is transferred into the load as well, but the time of transferring (τ_i)

$$\tau_L \approx \frac{L_3}{R_{l_3}}, \qquad (6.36)$$

may be many times longer than the time of capacitor C_4 (τ_l) discharge into the load resistor R_l

$$\tau_l = C_4 \cdot R_l . \qquad (6.37)$$

If SAS turn on time (τ_S) is many times less than discharge time ($\tau_S << \tau_l$) the loss of the load voltage ($\Delta U = U_3 - U_l$) in the first approximation is

$$\Delta U \approx U_m \cdot \frac{\tau_S}{\tau_l} .,$$

where U_l is the load peak voltage.

For the energy losses during turn on time (Q_{CS}) , assuming linear approximation of the SAS voltage drop, we have

$$Q_{CS} \leq \frac{\tau_S Q_{C_4}}{3\tau_l}, \qquad (6.38)$$

where Q_{C4} is the energy stored in C_4. For the losses (Q_{Son}) due to residual voltage drop determined by (6.14), we have

$$Q_{Son} \approx Q_{C_4} \cdot \frac{E\alpha}{E_S}. \qquad (6.39)$$

When total capacitor C_4 is represented by only nonlinear DSRD (D_2) capacitance, the discharge of the diode into the load is accompanied by losses in neutral region of the diode. During the period of charging of the diode from the inductor L_4 the current density is chosen to be saturated $j_s \approx qN_dV_s$. During the period of discharging the current may be many times

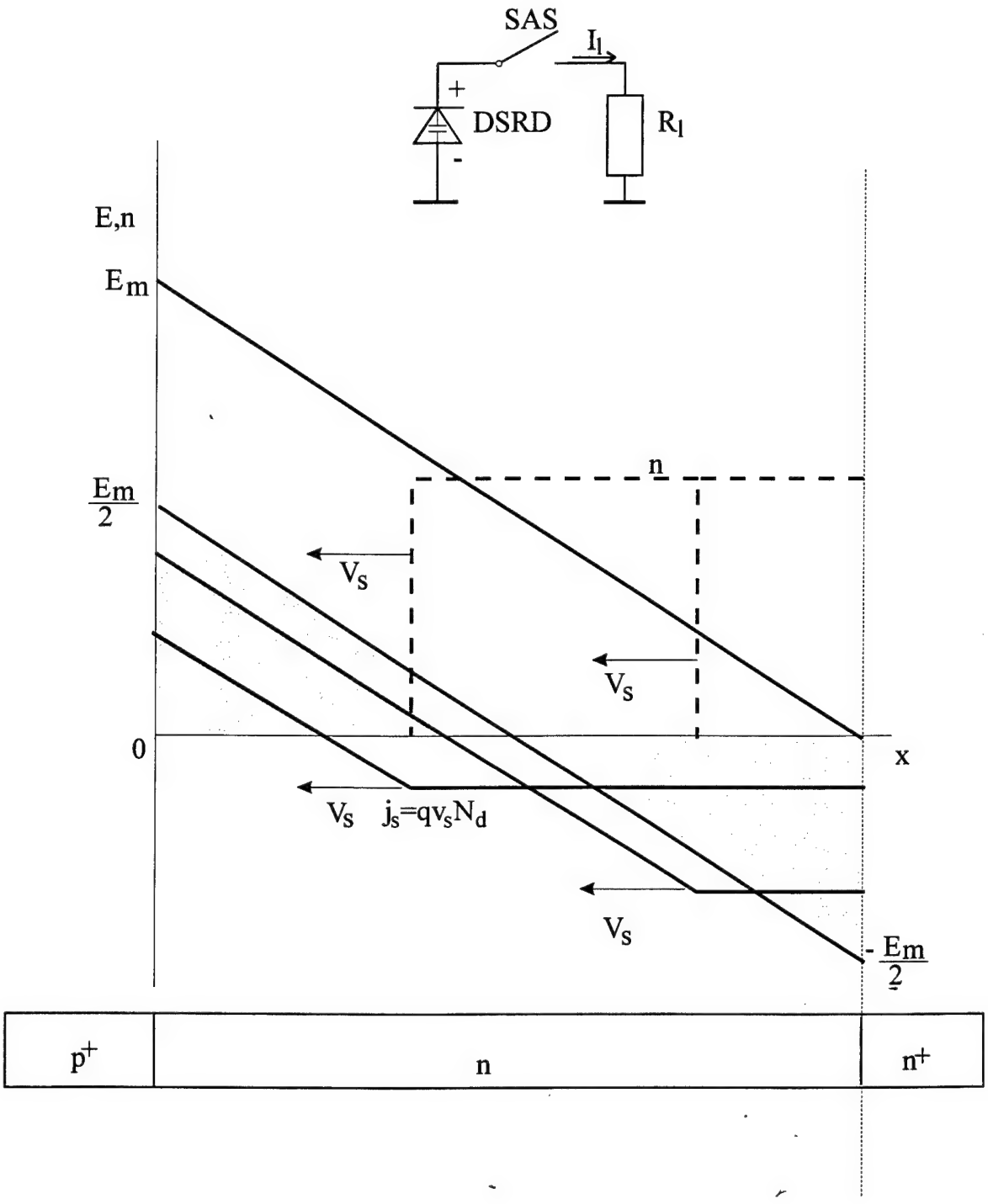

SAS
I_l
+
DSRD
-
R_l
E,n
E_m
$\frac{E_m}{2}$
n
V_s
V_s
0
x
V_s $j_s=qv_sN_d$
V_s
$-\frac{E_m}{2}$
p^+
n
n^+

Fig. 6.4

more than that during charging process. The process of discharge of a diode capacitance is different from those of usual linear capacitor.

Let us consider (fig. 6.4) the discharge of DSRD into the low ohmic resistor R_l after the fast switch SAS turns on. Let the turn on time of SAS (τ_{on}) and time of discharge $\tau_l \approx R_l \cdot C$ (where $C \approx \frac{S\varepsilon}{W}$) be many times less than the time of the flight of electrons through the n-layer (τ_S) $\tau_{on} << \tau_S = \frac{W_n}{V_s}$, $\tau_l << \tau_S$. It was mentioned that in DSRD the space charge region under maximum voltage overlaps n-layer (fig. 6.4). When the diode voltage drops to zero value, the space charge distribution is not changed $\left(\frac{dE}{dt} = \frac{qN}{\varepsilon}\right)$ because of short time of the discharge,. As may be seen from fig. 6.4, SCR is divided into two parts: positive field region and negative field region. The charge passed through the load (Q_R) during the time of the voltage drop is

$$Q_R = \int Idt = S\varepsilon\Delta E = \frac{\varepsilon E_\alpha S}{2}. \qquad (6.40)$$

For the energy stored in the diode initially before discharge (Q_0) and after (Q_p) we have

$$Q_0 = \frac{S\varepsilon E_\alpha^2 \cdot W}{6}, \quad Q_p = \frac{S\varepsilon E_\alpha^2 W}{24} \qquad (6.41)$$

The expressions (6.41) show that only 3/4 of the stored energy can be transferred to the load during short time period $\tau_p << \tau_S$. Then the process of neutralization begins: electrons come from n+ layer into the negative field region and compensate the space charge. The neutral region of compensated charge propagates into the n-layer with the saturated velocity of electrons. In the neutral region the current is the sum of the displacement current and conductivity current $j = \varepsilon\frac{dE_n}{dt} + qN_dV_s$. In the collapsing SCR only the displacement current exists $j = \varepsilon\frac{dE_{SCR}}{dt}$. From the conditions of continuity of current and zero voltage drop (see fig. 6.4) we have

$$j_r = j_s/2, \; E_{(x=0)} = \frac{qN_d}{\varepsilon 2}(W - V_s t). \qquad (6.42)$$

The expression (6.42) shows that the residual current j_r flows during the flight time $\approx \frac{W}{V_s}$, and the energy of the field remained after the fast discharge dissipates into the heat during this time.

The additional linear capacitor C_4 (fig. 6.1) provides better efficiency, but the payment is the slower voltage rise on DSRD and SAS. When energy stored in C_4 is equal to the energy stored in DSRD and the loss of energy in DSRD decrease down to 1/8 or 12,5%, the voltage rise rate is two times less. This decrease of the rate can worsen the SAS performance (see the part 5.3.1).

6.2.5 Examples for 100KV output (conclusion to the part 6.2)

As was shown above the efficiency of power compression strongly depends on the actual pulser design. The change of some parameter could increase efficiency of one cell and decrease efficiency of other cell. To illustrate the procedure of the estimations of efficiency made above we will consider two example of pulser with the highest output power which can be reached by use of one device stack at the output.

At the first we will consider the pulser with 1,5 ns FWHM output. No SASs will be used in this case (the dotted line load at fig. 6.1.

The expression (6.8) and (6.10) show that the load resistance to obtain maximum power is

$$R_l \approx \frac{U_0}{I_m} \approx \frac{C_W}{2C^2\varepsilon}, \qquad (6.43)$$

which yields for oil filling and silicon device $R_l \approx 75 \div 150$ Ohm, which nearly coincides with standard transmission line impedance.

The limit to voltage (6.8) denotes that it is the stack inductivity, that limits the pulse front. For practice, the inductivity influence must be many times smaller than the turn on time. Assuming the influence less than 10%, for the voltage limit from (6.7) (6.8) we have that $U_0 \approx 100$ kV. In accordance with (6.43) the load impedance is $R_l \approx 75$ Ohm and the maximum current in the load is $I_{lm} \approx 1300$ A. In accordance with (6.31) the braked by DSRD current is $I_m \approx 1800$ A. Output power is $P_m \approx 180$ MW.

For FWHM ≈ 1,5 ns, from (6.30) we have $C_4 \approx 4$ pf.

It is evident that the turn off time $\tau_{off} \approx \frac{W}{V_s}$ should be less than the pulse FWHM. From this condition we have:

$$W_n \approx \tau_{off} \cdot V_s \approx 7 \cdot 10^{-3} cm$$

A number of diodes in the stack we can get from (6.3), assuming $E_\alpha \approx 2 \cdot 10^5 V/cm$.

$$N \approx 140$$

and from (6.6) the minimum length of the stack is

$$W_S > N \cdot W_n = 1 \text{ cm}.$$

The actual length including p^+, n^+ layers and soldering interfaces is many times more. Current inexpensive technology can not produce diode wafers thinner than $(1,5\text{-}2)\cdot 10^{-2}$ cm. So the expected length of 100 kV stack is $W \gtrsim 2$ cm.

From (6.35) we have

$$S = \frac{3C_4 W_S}{4s} \approx 3cm^2,$$

where W_S is equal to total SCR width 1 cm.

The expression (6.29) gives the value for the inductor L_3:

$$L_3 \approx 4C_4 \cdot R^2_l \approx 90 \text{ nH}.$$

The condition of the total current transfer from L1, L_2 into L_3 gives:

$$L_1 = L_2 = 2L_3 = 180 \text{ nH}.$$

From (6.22) assuming $\tau_t \approx 10$ ns we have the C_2 capacitor value

$$C_3 = \frac{2\tau_t^2}{L_3\pi^2} = 220\, pf.$$

Let us estimate the losses in the parts considered above.

The relative energy Q_{S2} lost in D_2 during "on" state is determined by (6.33)

$$\frac{U_{S_2}}{U_m} \lesssim 0{,}03,$$

(where $\tau_t = 10$ ns, $\tau_t \approx 0{,}7$ ns were used) and may be evaluated from (6.22) and (6.33)

$$Q_{S_2} \lesssim U_{S_2} \int_0^{\tau_t} I_3 dt \approx U_{S_2} I_m \cdot \tau_t \left(\frac{1}{2} - \frac{1}{\pi}\right) \approx 4 \cdot 10^{-2} J \qquad (6.44)$$

The energy stored in L_3 when $I_m = 1800$ A is

$$Q_{L_3} = \frac{L_3 I_m^2}{2} \approx 0{,}15 J.$$

The total efficiency of L_3C_4 circuit is

$$\eta_3 \approx \frac{Q_{L_3} - Q_{S_2}}{Q_{L_3}} \approx 0{,}73.$$

The losses should be compensated by the current increase up to value $I_m' = 2200$ A, which makes the D_2 current approximately two times more than the load current.

The maximum voltage on D_1 is derived from (6.22)

$$U_{2m} = \frac{I_m}{C_3 \varpi_2} = \frac{I_m \tau_t}{C_3 \pi} = 32kV.$$

Assuming halfperiod of L_1C_1 circuit T = 200 ns, we get the values for capacitors C_1 and C_2.

$$C_1 = C_2 = \frac{T^2}{\pi^2 \cdot L_1} = 20\ nf.$$

The initial voltages on C_1 and C_2 (U_0) are

$$U_0 = \frac{I_m \rho}{2} = \frac{I_m}{2}\sqrt{\frac{L_1}{C_1}} \simeq 3{,}3\ \text{kV}.$$

The voltage drop on D_1 during on state in accordance with (6.25) is

$$U_{S1} < 150\ \text{V},$$

if $\tau_{off} = 5$ ns and $W_n = 10^{-2}$ cm are assumed.

The energy losses, during on state (for sine like current) in D_1 are derived from (6.25) for reverse current:

$$Q_{S_1} \lesssim 2U_{S_1} \int_0^{T/2} I_1 dt = U_{S_{1-m}} \int_0^{T/2} \sin \varpi t\, dt = \frac{U_{S_1} I_m \cdot T}{\pi} = 0{,}012\ J.$$

For pumping current the losses are nearly the same ($\approx\sqrt{2}$ times less) $Q_p \approx 0{,}08$ J. So total losses on D_1 are $Q_{D1} \approx 0{,}02$ J. The comparison of these losses with the energy stored in the inductor L_3 shows that correction for the currents in L_1 and L_2 inductors are small and corrected values : $I_1 = I_2 = 1200$ A.

The assumed value for τ_{off} in accordance with (6.24) gives:

$$j_1 = \frac{\varepsilon E_\alpha}{\tau_{off}} \approx 40\ A/cm^2$$

$$S_1 = \frac{I_m}{j_1} = 60\ cm^2,$$

where S_1 is the area of $DSRD_1$ stack.

The number of diodes in the stack is derived from (6.3): $N_1 = 3$.

The energy returned to the capacitors C_1 and C_2 is (see (6.28))

$$\frac{Q_r}{Q_0} \approx \frac{4\tau_f^2 \cdot \pi^2}{3T^2} \approx 0{,}13$$

These losses should be compensated by additional increase of the currents up to the values

$$I_1 = I_2 \approx 1300\ \text{A}.$$

To compensate total losses, considered above, the energy stored in the inductors L_1 and L_2 must be

$$Q_L \approx 0{,}3\ \text{J},$$

The load energy is

$$Q_R \approx 0{,}15\ \text{J},$$

So the total efficiency is 50%.

Let us consider the requirements to the primary closing switches S_1 and S_2 :

1. blocking voltage: $U_0 > 3{,}5$ kV
2. maximum current: $I_1 \approx 1300$ A
3. pulse length: 200 ns for the S_1, 100 ns for S_2.

The expression (6.19) shows, that the requirements of low losses (< 10% for Q_C and < 10% for Q_S) may be satisfied when:

1. turn on time τ_f is less than $\frac{1}{\varpi}\sqrt{\frac{3Q_C}{Q_0}} \approx 34\ ns$
2. on state resistance R_S is less than $\frac{\rho \cdot Q_S}{\pi Q_0} \lesssim 0.1$ Ohm.
3. dI/dt factor is many times more than 40 A/ns.

Some semiconductor power devices have been considered in the part 3. No devices which could satisfy to this requirements, are known to be produced on mass scale.

It is possible to increase switching voltage by connecting many devices in series (for example 4 devices, each rated at 1 kV). Every type of devices, considered in the part 3, may be connected in series.

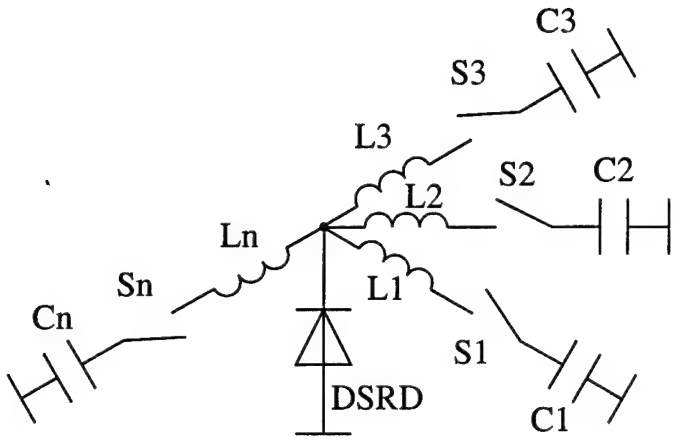


Fig.6.5

One of the advantages of the power compressing network (fig. 6.1) is that $C_1L_1S_1$ (and $C_2L_2S_2$ as well) circuits may be subdivided into many connected in parallel CLS circuits as shown in fig. 6.5. The current form in the circuits is determined by only LC parameters, if the Q-factor is high enough ($\gtrsim$10). It is well known fact from the practice that LC parameters are controlled with good precision, better than 1%. Their time stability is better than 10^{-4}. The circuit fig. 6.5 due to small voltage drop on DSRD is nearly ideal summing system and a large number (no limits is seen) of low current circuits can provide synchronously high currents (pumping and reversed) for DSRD .

Let us check the possibility of the use the different types of devices discussed in the part 3 as a primary switch. If Q-fqctor is large (Q>6), the relative energy losses due to inequality of the L_1 L_2 currents (see (6.23)) are small- several percents. For $Q \approx 2$ the losses may be > 20%!

1. 1 kV rated FET IRFBG30 provides only 8 A, when the Q factor is 6 and losses are $\approx$ 20%. A number of parallel LC circuits is 325, total number of FET is 1300 (4 units is in series).

2. Bipolar transistors give nearly the same results - total number is $\approx$ 1000.

3. 1 kV rated thyristors provide 50 ÷ 100 A each, when the Q factor is 6. Total number of thyristors is 200-100 (4 units in series).

4. New devices - dynistors are the most promising switches. 1-2 kV rated dynistors provides up to 500-700 A with good factor ($Q_f \approx 6$). Total number is 16, which is many times less than the numbers evaluated above.

With good Q-factor ($Q \approx 6$) of the primary switches overall efficiency $\approx$ 40% is possible for 100 kV, 1,5 ns pulser.

It is necessary to stress that the most critical part of the pulser under consideration is not new fast devices, but slow primary switches.

The pulser considered above may be easily scaled down to lower voltages by simple decrease of stacks size (both - length and diameter). The pulser after the scaling may be used as the initial stage before the last subnanosecond compressing cell based on SAS (fig. 6.1). If the load resistor connected in parallel with DSRD2 is removed, the DSRD2 voltage becomes higher than estimated above.

In the part 6.2.4 was shown that the turn on voltage of SAS must be $\approx$ 0,85 of that õn D_2. Assuming 100 kV output, permitted by (6.8), for SAS area (6.13) and capacitance (6.35) we have

$$S \approx 10^{-1}\ \text{cm}^2, \quad C_{SAS} \approx 0{,}25\ \text{pf}.$$

In accordance with (6.34) we have that the current I_3 charging D_2 capacitance up to needed 125 kV is

$$I_3 \approx \frac{U_3}{\sqrt{\frac{L_3}{C_4}}} \approx 850\,A.$$

The estimation of the current shows that all preceding compression cells could be scaled down slightly (scale factor is ~ 0,7) by simple decrease of initial voltage at capacitors C_1 and C_2.

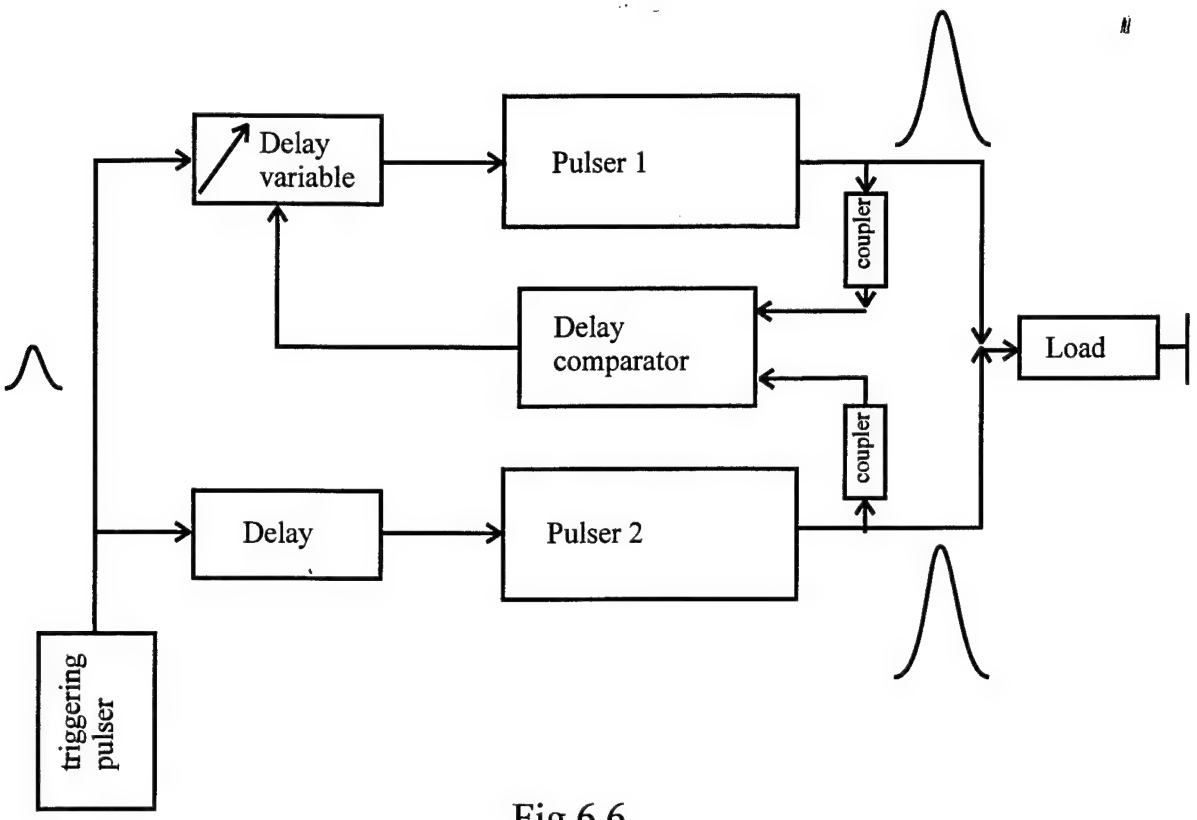

Delay variable
Pulser 1
coupler
Delay comparator
Load
coupler
Delay
Pulser 2
triggering pulser

Fig.6.6

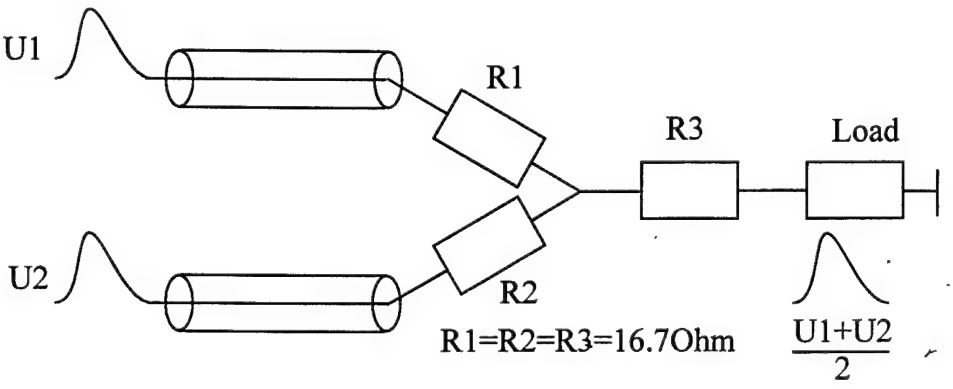

U1
R1
R3
Load
U2
R2
R1=R2=R3=16.7Ohm
U1+U2
2

Fig.6.7

It should be noted, that the inductance of the D_2 stack is too large for 100ps front generation. So the capacitor C_4 must be separated into two parts: low inductive additional capacitor and DSRD having smaller area.

In accordance with (6.37) for 50 Ohm load from (6.37) we have for $C_4 = 4$ pf

$$\tau_1 = 200 \text{ p.}$$

The losses in the subnanosecond cell are:

1. The losses due to nonlinearity of D_2

$$Q_1 \approx 0{,}25\ Q_{C4}$$

2. The losses due to dU/dt requirement

$$Q_2 \approx 0{,}25\ Q_{C4}$$

3. The commutation losses (6.38)

$$Q_3 \lesssim 0{,}5\ Q_{C_4}.$$

The efficiency of the last cell, evaluated from the considered above losses is ≈ 40%. The overall efficiency of all stages is ≈ 20%.

In conclusion we will remark that the consideration shows the feasibility of the design of the pulser with 100 kV output into 50 Ohm with 100 ps rise, 200 ps decay times and efficiency ≈ 20%.

6.3 Time stability and synchronization

As was shown before, the most efficient method of pulse generation is power compression using several compressing stages. In this case the total delay τ_d of the output short power pulse from the first pulse triggering the primary switch (a thyristor or power transistor) may be as high as several hundreds nanoseconds (100-400 ns). The delay τ_d, as a rule, depends on the output power (the more power, the more delay) due to the fact that more powerful primary switches (thyristors or transistors) are slower.

In many applications, for example when many pulsers work at the same load to increase power, the stability of the delay is very important factor. For the mentioned case of the power summation of pulses having subnanosecond fronts, the absolute instability ($\delta\tau$) must be less than the front $\delta\tau < 100$ ps and relative unstability (d) must be as low as $\delta \approx \frac{\delta\tau}{\tau_d} \leq 10^{-3}$. Delay instability consists of two components: slow (drift) when change of delay period is minutes and fast (jitter) - the change of delay from one pulse to the next. In pulsers slow drift may be as large as nanoseconds and even tens nanoseconds for the delay of hundreds ns. The last large value is due to the "warm up". After "warming up" during 20÷30 minutes the drift goes down to nano and subnanosecond range in tens minutes, depending on the temperature stability of the environment, mechanical vibrations and so on. For summing up of many pulsers outputs into the single load or the change of an radiating antenna pattern slow drift may be compensated by back feed loop. For example at the pulse repetition frequency (PRF) 1 KHz it is possible to use one pulse per a second (1 pulse from 1000) to adjust the change in time position of the output pulses by change of the delay of triggering pulses (see fig. 6.6).

The fast and random changes from one pulse to the next can not be compensated and we will consider the jitter in more details.

The main sources of instability of the delay for the most used pulse compressing circuit (fig. 6.1) may be represented as:

1. instability of the delay between two circuits triggering the primary switches S_1 and S_2.

2. instability of the L_1, C_1 and L_2, C_2 circuits and, as a result, instability of half period oscillation of the LC circuits.

3. instability of the "turn -on" delays of the primary switches S_1 and S_2 in respect to their triggering.

4. instability of delays of DSRDs D1, D2 turn-off moments .

5. instability of the high voltage sources charging the primary storage capacitors C_1, C_2 peaking capacitor C_4 and blocking C_5.

Let us consider the listed above causes of instability in more closely.

1. There are well known approaches to decrease the instability (jitter) of driving circuits. For example - direct counting of high frequency highly stabilized oscillations, use of miniature LC delay lines and so on. The applications of such methods in samplescopes and digital scopes allows to get jitter as small as tens picosecond at the delays as long as 100÷200 ns.

2. It is well known fact that relative long time (hours) stability of LC circuits of high frequency generators (at tens MHz) may be as high as $10^{-3}\div10^{-4}$. The frequency drift for short time period (less than 10^{-2} second), corresponding to our pulse repetition rate is far less. That means, jitter of half period (δT_{LC}) determined by LC stability is less than $\delta T_{LC} < 10^{-4} \cdot T_{LC} \approx 100$ ps, where T_{LC} is LC's half period.

3. The primary switches S_1 and S_2 are power thyristor or transistors (FET or bipolar). The total delay of this devices depends on the ratio of the gate current to the maximum collector current. The more the ratio, the less the delay. As it was shown in section 3, in the case of switching on LC circuits two different cases are possible:

a. current rise rate (dI/dt) is limited by maximum value of the switch dI_m/dt factor.

b. current rise rate is limited by LC circuit impedance, that is the condition

$$\frac{dI_m}{dt} >> \frac{U_o}{L}, \tag{6.45}$$

(where U_o is the C_1, C_2 charging voltage, L is the inductance of L_1 or L_2 inductors) is fulfilled.

As was mentioned above, for the second case (6.45) the transfer of energy from the capacitor to the inductor is more efficient: the energy loss is less (6.17) and so influence of the turn on processes on current rise is low. In this case the delay of the device turn on is determined by the collector voltage drop, but not the current rise up to some level (usually 0.9 form maximum).

As the first approximation, the delay time (for bipolar devices) is determined by diffusion time across the gated base layer (τ_{dif}) and the time of collector space charge region discharge (τ_{SCR})

$$\tau_{diff} \approx \frac{W_g^2}{2D},\ \tau_{SCR} \approx \frac{1}{j_g} \cdot \sqrt{2\varepsilon U_0 q N_d} \tag{6.46}$$

where j_g - triggering current density at the collector

U_o - collector voltage

N_d - collector layer doping

W_g - base layer width

D - carriers diffusivity.

τ_{SCR} may be made as low as several nanoseconds even for high voltage and power devices by using of fast power triggering . The minimal discharge time is limited by the time of flight of carriers through the space charge region. For high voltage devices the time is near 1 ns per 1 kV blocking voltage.

The estimated τ_{dif} for 1kV rating power bipolar transistors or thyristors is equal to tens nanoseconds. This value corresponds to the unity current gain: after $\tau=\tau_{dif}$ the collector

current discharging the SCR capacitance reaches the triggering current value, provided that the LC circuit's current during this delay time is small due to dI/dt limitation ($dI_{LC}/dt \le \frac{U_o}{L}$ and $I_{LC} \le \frac{U_o}{L} \cdot \tau_{dif}$.

It is possible to decrease the time during which the collector current discharging the SCR capacitance reaches the needed level by increase of the triggering (gate) current. But this way is not very efficient: the delay is decreased sublinearly with the current increase.

The most efficient way is to decrease the base layer width and diffusion time τ_{dif}.

The extreme position of this approach nearly overlaps with the case of field effect transistor: all triggering (input) current is spent to charge the gate capacitance which is more than collector one.

So, it is possible to achieve delay of power switch (τ_d) as short as tens and even less than ten nanoseconds.

It is well known that the carriers mobility and diffusivity in semiconductor strongly depend on the temperature. When the temperature changes in the range ($0^0 \div 40^0$C), the mobility change is as high as tens percents. But the temperature causes relatively slow change - drift, but not the jitter considered here.

Let us consider the influence of random motion of carriers on the jitter. The times of momentum relaxation in semiconductors are very small $\approx 10^{-13}$ sec.. The number of carriers in the volume of power device is large $N > 10^{12}$. That is the relative fluctuation of the number (δN) is

$$\delta N \sim \frac{1}{\sqrt{N}} < 10^{-6}.$$

The relative fast fluctuation of the current has approximately the same value. It is evident that the relative fluctuations of the time intervals during which the charge transferred by the current reaches some definite level have approximately the same value $\delta\tau_d \approx \delta Q \approx \delta I \approx \delta$. So the jitter connected with carriers fluctuation is $\delta T_d \approx \delta\tau_d \times \tau_d$ and far less than 10^{-12} sec. It should be noted that the same estimations are valid for processes of accumulation and dispersal of plasma in DSRD.

4. The time of the current break by DSRD (T_b) is determined by the following condition: the charge (number of carriers) stored in the diode during the pumping phase (Q_+) by forward current I_+ is equal to the charge (Q_-) pulled out by the reverse current (I-)

$$Q_+ = \int_0^{T/2} I_+ dt = Q_- = \int_0^{T_b} I_- dt \qquad (6.47)$$

The expression (6.47) is valid if the carriers losses are small. As was mentioned before, DSRD have to have as large life times (τ_p) of carriers as possible. Modern technology allows to make the times larger than 100 μs and recombination losses less than $<<10^{-3}$ for half period of ~100 ns.

The second possible source of losses is carriers leakage (through p^+n and nn^+ junctions in asymmetrical) and near contact p^+p and nn^+ junctions in quasisymmetrical diodes. The leakage may be described by injection coefficients of the p-n junctions (γ).

The shift of the break point ΔT_b due to charge losses is proportional to the charge losses:

$$\frac{\Delta T_b}{T/2} \sim \frac{T}{2\tau_p} + \frac{1-}{\gamma} \qquad (6.48)$$

In the case of $\gamma \approx 0{,}999$, $\tau_p > 100$ μs, $T/2 \approx 100$ ns, (6.48) yields $\Delta T_b < 0.1$ ns.

It is evident that fast fluctuations (jitter) between two pulses δT_β are determined by fluctuations of τ_p and γ and must be far less than the ΔT_b evaluated above.

5. instability of voltage source δU could affect the jitter in the variety of manners.

5a. When both capacitors C_1 and C_2 are charged from the single source, the change of the source voltage changes both Q_+ and Q_- preserving their ratio and, therefore, the position of the braking point.

Therefore T_b is not sensitive to δU

5b.The delay of SAS switching on (τ_{SAS}) is not more than several nanoseconds (<3 ns usually) from the start of fast voltage rise. It is well known that devices, which use impact ionization, such as avalanche photodiodes and so on, are subjected to strong current instabilities. The main source of the devices current fluctuations is the strong relative fluctuation of the small number of primary carriers initializing impact ionization.

It was shown in the section 5.2, that in SAS the number of initial carriers starting impact ionization in neutral region is large (majority carriers - electrons). Their concentration is ~10^{14} cm^{-3} and their number $N>10^{11}$ leads to the relative fluctuations of the number $\frac{\delta N}{N} \approx \frac{1}{\sqrt{N}} < 10^{-5}$. A number of holes generated by ionization in neutral region (P_0) is $P_0 \approx 10^6$ and their relative fluctuation is $\frac{\delta P_0}{P_0} \approx 10^{-3}$ far larger, than that of electrons. The evaluation of delay time fluctuation due to that factor is

$$\delta\tau_{SAS} \approx \tau_{SAS} \cdot \frac{\delta P_0}{P_0} < 3 \text{ ps}$$

rather small value.

As it was shown, the power supply instability does not bring the instability of DSRD's current break point. But the supply instability could cause strong instability of the closing switches delay - SAS and primary ones S_1 and S_2.

In primary switches, as it follows from (6.46), the instability of turn on delay ($\delta\tau_d$) is proportional to the voltage instability (δU_0):

$$\delta\tau_d \approx \tau_d \cdot \frac{\delta U_0}{2U_0} . \qquad (6.49)$$

The same expression taking into an account the new value of the delay is valid for the case of SAS.

It follows from (6.49) that instability of power supply (providing jitter less than 10ps at delay $\tau_d \approx 10$ ns) must be better than 10^{-4} ($\frac{\delta U_0}{U_0} \leq 10^{-}$).

Total fluctuation of output pulse delay ($\delta\tau_{out}$) may be represented as

$$\delta\tau_{out} \approx \sqrt{\sum_i \delta\tau_{d_i}} , \qquad (6.50)$$

where $\delta\tau_{di}$ are partial fluctuations.

Estimation for (6.50) shows that it is possible to get total fluctuation (jitter) as small as tens picoseconds.

It should be remembered that slow drift, which an orders of magnitude larger (nanoseconds) may be compensated by using of the feedback loop.

The same (6.50) expression determines the average dispersion ($\delta\tau_m$) of jitter for the case of many pulsers (m) working into the same load:

$$\delta\tau_m \approx \delta\tau_d \cdot \sqrt{m} \qquad (6.51)$$

For example, in the case of 10 identical pulsers working at the single load the total dispersion is $\delta\tau_m \approx 30$ ps for 10 ps jitter of single unit. So, the synchronization of 10 pulsers with 100 ps fronts in feasible.

The evaluations made above, were checked at the example of 2 identical pulsers assembled on a large printed board with 7 kV output into 50 Ohm. Two outputs were fed into the load resistance through star-like matching circuit (fig. 6.7) having 6 db attenuation. Each pulse has ≈ 200 ps front and ~1 ns decay at PRF ≈ 500 Hz. After adjustment of the delay between outputs for zero level, the load pulse front and voltage were equal to the output of each

pulser. Then during half an hour due to slow drift of the delay the load pulses amplitude decreased slightly ~20% and two "peaks" appeared on the load pulse.

The long time of pulses divergence makes feasible the automatic control of synchronization mentioned above.

6.4 Average power and pulse repetition frequency (PRF)

6.4.1 General consideration

6.4.2

At the chapter 6.1. peak power and electrical efficiency have been considered. It is evident that output average power (P_a) is the product of the pulse energy (Q_R) and PRF (f_P)

$$P_a \approx Q_R \cdot f_P \qquad (6.52)$$

The limitation on the average power may be determined by (6.52) when there are limitations on the pulse energy and/or PRF. It is possible as well, that average power is limited by heat sink and overheating of some critical parts, as a rule, semiconductor devices. In this case the relation (6.52) should be used to determine and limit PRF when average power and pulser energy are given, or to determine pulse energy when average power and PRF are given. From mentioned cases it follows that for general case, the average power is determined by very complicated interplay of different factors and may be clearly stated only for actual pulser design.

Therefore here at first we will consider the thermal (heat-sink) limitation on average power. Then after considering of non thermal limitation PRF, it will become possible to estimate average power for actual pulser as minimal from two thermal and nonthermal limitations.

6.4.3 Thermal limitations

It is evident that thermal limitations on devices in the chain of pulse compressing cells may differ strongly. The overheating of primary switches (thyristors, transistors) is well known problem and have been much studied elsewhere. We will consider here only DSRD and SAS heating.

6.4.3.1 DSRD heating

The thermal diffusitivity D_T in semiconductors are small (for Si $D_T \approx 0.7$ cm^2/c) and the heat (due to dissipated energy) can not be pulled out of the thick semiconductor bulk into the external heat sink during the short time of heating. Diffusion time of heat (T_W) through the typical n-layer of thickness $W_n = 10^{-2}$ cm is

$$\tau_W \approx \frac{W_n^2}{D_T} \approx 100\ \mu s,$$

the time that is much longer than the halfperiod time during which the energy dissipates in the layer.

The temperature increase in the n-layer (ΔT_p) during pumping time (τ_p) may be estimated from (2.26), (6.20), (6.24), (6.26), (6.27)

$$\Delta T_p = \frac{Q_\tau}{VC_\tau} \approx \frac{U_S j_+ \cdot T_+}{2 \cdot W_n \cdot C_\tau} \le \frac{20 U_2 \sqrt{D}\, T\varepsilon}{3\tau_{off} \sqrt{\tau_{off} \mu_n \cdot \mu_p}\, W_n} \qquad (6.53)$$

where C_τ - specific thermal capacity, $V = S \cdot W_n$ - the n-layer volume, T - half period (see (6.25)) of the first or the second stage compressing cell (6.27) depending on the case under consideration.

In the case of $\tau_{off} \approx 5$ ns, T = 200 ns, $U_2 = 1$ kV, $W_n \approx 10^{-2}$ cm, $C_\tau \approx 2$ J/cm^{3}K expression (6.53) yields $\Delta T_p < 10^{-2}$ ^{o} C- very small value. The other heating sources (reverse current and transient switch off process (6.32)) only slightly increase the estimated value. That means very small deviation of the peak temperature from average after "warming up" time. After "warming up" the average temperature is determined by the balance between incoming heat ($Q_\tau \cdot f_P$) and outcoming into the heat sink (P_{out}).

Assuming the ideal heat sink with zero thermal resistivity, we get the thermal flow density

$$\widetilde{P}_{out} \approx \lambda \frac{\Delta T}{W}, \tag{6.54}$$

where λ is the thermal resistivity of semiconductor, ΔT - temperature drop on the device bulk.

It should be noted that total thermal flow is proportional to the area of the device and so is the total DSRD current. Therefore the heat balance does not depend on the area of the device and the total device current.

For a device with only one p-n junction and even one side cooling from (6.54), using the same as for (6.53) approach, we get

$$\Delta T = \frac{W \cdot Q_\tau \cdot f_P}{\lambda} \text{ or } f_P = \frac{\Delta T \lambda}{W \cdot Q} \tag{6.55}$$

$$\frac{Q_\tau}{S} \lesssim \frac{U_S \cdot j_+ \cdot T}{\varepsilon} \approx \frac{5 U_2 \sqrt{D}\, T \cdot \varepsilon}{\tau_{off} \sqrt{\tau_{off} \cdot \mu_n \cdot \mu_p}} \tag{6.56}$$

where S is the device area.

For the case, considered above, (6.56) yields $\frac{Q_\tau}{S} \lesssim 10^{-3}$ J/cm^2 - very small value.

The increase of DSRD temperature decreases carriers mobility and worsens turn off time. Experiments showed that heating up to > 150°C increases turn off time by ~ 30%. Assuming the maximum temperature 150°C we get from (6.54) the maximum possible cooling capacity $P_{out} \approx 10$ KW/cm^2. The maximum possible PRF from (6.55) is $f_P > 10^8$ Hz. Of course the actual PRF (f_{ac}) for the example is limited by $\frac{3L_1C_1}{2}$ period time $f_{ac} < 3 \cdot 10^6$ Hz. Consideration made above has shown that thin low voltage DSRD used at the first compressing stage potentially has no heat limitation. In actual design the influence of external thermal contact resistance (thin soldering interface, tungsten or molybdenum thermal expansion compensator and so on) may be large and may increase the thermal resistance tens times. Nevertheless the main limitations still will be with LC period time.

DSRD used in the second stage of pulse compression (see Fig. 6.1.) have far less turn off times ($\tau_{off} < 1$ ns) and higher thermal losses may be expected form (6.56), but the "pumping" time and the length of reverse current period are shorter in nearly the same proportion as τ_{off}. So, it may be expected that the density of thermal losses remains nearly the same as in the case of the first stages and so does the limitation on PRF. As a rule, due to smaller charge stored in the second stage DSRD, the condition $t < t_u$ is fulfilled in (4.21) and the voltage drop is determined by the first expression from (4.21) for the pumping case and by (4.28) for the reverse current case. As it was mentioned before, (4.28) is a "mirror" reflection of (4.21) in respect to the moment when the current changes polarity, and the losses may be calculated by any of them for both cases, taking into an account the corresponding current pulse length. The contribution of the DSRD turn-on state losses at the second stage into the total pulser efficiency is small in respect to other losses and the loses have not been considered in detail in the part 6.2.3. But the losses may contribute significantly into the incoming thermal heat of DSRD and will be considered here lately.

In the case of high voltage stacks of many DSRD the condition of balance between incoming and outcoming heat changes drastically. Incoming heat flow (6.56) is increased

proportionally to the stack voltage. Outcoming heat flow (6.54) is decreased proportionally to the thickness of the stack and voltage.

For the stack (6.54) may be rewritten as

$$\widetilde{P}_{out} \lesssim \lambda \frac{\Delta T \cdot E_\alpha}{U_{st}},$$

where U_{st} - the stack voltage.

For the case of $U_{st} \approx 100$ kV and $E_\alpha \approx 100$ KV/cm we have: $\widetilde{P}_{out} \lesssim 150$ W/cm^2.

Again we should remember that the estimations are very approximate and we do not take into account heat resistance of interfaces between p-n junctions and heat sink, which may be large.

It was shown that maximum working area of DRSD is limited by (6.9) and so is limited the heat flow P_{max} coming into the heat sink

$$P_{\max} \lesssim \widetilde{P}_{out} \cdot S \tag{6.57}$$

From (6.9) in the case of $\tau_f \approx 1$ ns follows $S \lesssim 10^2$ cm^2 and $P_{max} \lesssim 15$ KW.

The output (at the load) average power P_l is

$$P_l = \frac{P_{\max}}{1-\eta} \tag{6.58}$$

where η - is efficiency of the second stage DSRD.

The efficiency, as it was mentioned, is determined by the losses at "on state" and by the transient (commutation) losses (6.32).

When "pumped" charge Q_P is small

$$Q_P < \varepsilon S q W N_d \mu_n / \mu_p,$$

the voltage drop at the diode during "on state" is determined by (4.21) of (4.28), each of which (see the remark above) may be represented for the case of constant current

$$U_S = \frac{j_+}{q\mu_n N_d}\left(W - \frac{tj_+\mu_p}{\varepsilon q \mu_n N_d}\right) \tag{6.59}$$

The expression (6.59) shows that diode voltage is less than voltage drop on the not modulated (not enriched) n-layer having resistance $R_S = \frac{W_n}{Sq\mu_n N_d}$. The energy loss during pumping will be only less than "Q_P"

$$Q_P = \frac{R_S I_+^2 \pi \tau_+}{4} \tag{6.60}$$

where τ_+ is the length of the pumping current pulse, I_+ - the current amplitude.

Taking into an account the equality of pumped and extracted charge we have from (6.60)

$$Q_P = Q_- \frac{\tau_-}{\tau_+} \tag{6.61}$$

$$Q_- = \frac{\pi R_S I_-^2 \tau_-}{4},$$

where Q_- is the energy loss at R_S during reverse current period, I_- - breaked current.

From expression (6.61) follows, that losses during pumping are less than during reverse current, if the pumping period is longer (usually this is the case). Actually the ratio τ_-/τ_+ may be as small as < 0.1 and the "pumping" losses in the second stage DSRD may be neglected. It may be shown that the same as (6.61) relation is valid in the case of strong modulation, because the carrier distribution depends on only injected charge, not the actual curve of the current versus time (see the part 4.3).

The current must be near to "saturated" value $j_- \approx j_S = qV_S N_d$ to get the shortest "turn off" time. In this case from (6.61) we have

$$Q_- = \frac{\pi E_S W_n I_- \tau_-}{4} = \frac{\pi U_m I_- \tau_- \cdot E_S}{2 \cdot E_\alpha}. \tag{6.62}$$

It may be shown that (6.62) is valid in the case of a diode stack as well.

In the case of bell like pulse (see the condition (6.29)), taking into an account (6.29), (6.31), (6.35) we have for the energy transferred into the load (Q_l)

$$Q_l \approx 8U_m I \tau_{off} \qquad (6.63)$$

$$Q_{\sim} = \frac{E_S}{E_\alpha}\frac{\pi\tau_{\sim}}{16\tau_{off}}Q_l$$

The expression (6.63) shows that pulse compression should not be large for good efficiency. If $\tau_{\sim}/\tau_{off} < 5$ the total losses in the DSRD stack are less than 10% ($\xi > 90\%$) (see (6.32) as well) and the average output power can reach (6.58) $P_l > 100$ KW. It should be emphasized that the second stage (output) high voltage DSRD is the "bottleneck" in the processes of pulse compression in respect to the average power.

In accordance with (6.63) and estimation for $S \approx 10^2$ cm^2 we have: $Q_l \approx 1$ J and maximum PRF, limited by heating, is $f_p \approx 150$ KHz.

6.4.3.2 SAS heating

As it was mentioned, DSRD are not sensitive to the temperature increase, their performance degradation is connected with mobility decrease only. Devices based on delayed ionization are more susceptible to the heating due to the fast (exponential) increase of leakage current. It has been shown in part 5.3 that large leakage prevents the field intensity increase above the breakdown threshold and generation of fast ionization wave with high density plasma "tail" (5.31). Experiments showed that fast switching under moderate $dU/dt \approx (1\text{-}2)\cdot 10^{12}$ V/s applied exists only at temperature under 100°C. The one pulse heating in SAS may be derived from (6.53) where income heat (Q_τ) has different value partually considered in part 5.4:

$$Q_\tau = Q_d + Q_{on} + Q_S$$

a. during delay

$$Q_d \lesssim j_S \frac{E_S \tau_d W S}{2} \qquad (6.64)$$

where τ_d - ionization delay $\left(\tau_d \approx \frac{U_m}{U'}\right)$;

b. during turn on process all energy stored in the electric field before the fast wave generation is spent on electron-hole pairs generation and in the end will be converted into the heat. This part of energy may be determined by the expression like (6.35). Additionally after the wave front (in the wave "tail") some energy is lost due to conductance current. This loss is determined by (6.64) like expression. For total losses we have

$$Q_{on} \lesssim \frac{2S\varepsilon U_m^2}{3W} + \frac{j_S E_S W S \tau_f}{2} \qquad (6.65)$$

where U_m - turn on voltage;

c. during turn on state, taking into account (6.14) we have

$$Q_S \approx j_m U_{on} \cdot \tau_p S < \frac{S\tau_p \cdot \varepsilon E_\alpha \cdot E_S W_n}{\tau_f}\left(\frac{E_\alpha}{b}\right) \qquad (6.66)$$

where ε_p - pulse length, U_{on} - sustained (on state) voltage in SAS.

The estimations on (6.64), (6.65), (6.66) and (6.53) for the case $\tau_d \approx 2$ ns, $\tau_f \approx 0{,}1$ ns, $\tau_p \approx 0{,}2$ ns, $U_m = 3{,}5$ kV, give: $\frac{Q_d}{S} \lesssim 10^{-5} J/cm^2$, $\frac{Q_{on}}{S} \lesssim 10^{-3} J/cm^2$, $\frac{Q_s}{S} \lesssim 10^{-5} J/cm^2$, $\Delta T \lesssim 10^{-1}$°

The main losses are turn on losses determined by loss of the energy stored in the electric field. The "one pulse" overheating is order of magnitude larger than in the case of DSRD, but still is small.

From (6.57), for the case of $\Delta T \approx 100$°C $U_{st} \approx 100$ kV and $E_\alpha \approx 3 \cdot 10^5 KV/cm$ (in SAS E_α is more than in DSRD), we get: $\tilde{P}_{out} \lesssim 300$ W/cm , slightly more than for DSRD due to higher E_α.

It should be stressed that in the case of SAS the maximum area is small ($S < 1$ cm^2 for $\tau_f \approx 100$ ps) and total allowed thermal income is less than 300 W.

Using estimation for the losses made above we get (from 6.55) maximum average PRF $f_p < 300$ KHz.

For maximum switched on current 2 kA (see (6.15)) and taking into an account (6.38) we have the energy that may be transferred into the load Q_l for the pulse width ≈ 0.2 ns: $Q_l \approx \frac{I_m \cdot U_m \tau_f}{3} + \frac{I_m U_m \tau_{dec}}{3} = 10^{-2}$ J, where τ_{dec} is decay time, $\eta \approx 90\%$ and average load power limited by heating is $P_l \lesssim 3$ KW.

6.4.4 Nonthermal limitations

6.4.4.3 . DSRD

In the preceding part it has been shown that potentially for DSRD thermal heating does not limit average power and repetition frequency in low voltage (~1 kV) pulsers.

The average power is limited by PRF, which in its turn is limited by the length of LC cycle for the circuit shown at fig. 6.1. Actually PRF is severely limited by the primary switches S_1, S_2, which have to operate at high PRF (megahertz). The time of discharge of storage capacitors C_1 and C_2 is 100 ÷ 300 ns, but the capacitors must be charged before the next cycle from power source. To get the charge time as short as the discharge period it is necessary to switch off very quickly the primary switches S_1, S_2. The turn off time should be as small as turn on time. Power FET transistors have short turn on and turn off times near 10-15 ns and may be used for the purpose. But at high PRF in constant mode (not burst) of operation the current switched by each of the transistor should be many times (order of magnitude) less than possible at low PRF (the subject has been discussed in the part 3.1). So the total number of FET used in pulser must be very high and the pulser design becomes very complicated.

There is another more efficient approach for high PRF pulsers design, based on the use of DSRT (Drift Step Recovery Transistors, considered in the part 4.4). This approach will be detailed in the next part.

6.4.4.4 . SAS

It has been shown that one the main condition for effective operation of SAS is low concentration of initial carriers and low conductance current leakage. After turn on process the volume of SAS (n-layer) is filled by high density ($10^{14} \div 10^{16}$ cm^{-3}) electron-hole plasma. It is evident that the next cycle of operation (the next pulse generation) is possible only when the plasma is completely dispersed (removed) and the remained carriers concentration drops down to very low level ≈ 10^6 cm^{-3}. The ratio of concentration in turn on and turn off states should be very large ≥ $10^9 \div 10^{10}$ times!

There are only the next ways of the plasma removal from SAS with diode structure p^+nn^+:

1. extraction of carriers from n-layer into the p^+, n^+ layers. The extracted carriers in their turn must be extracted from the layers into the contacts or disappear in them due to recombination. If the carriers remain in p^+, n^+ layers, during the next cycle the high displacement current in SCR of p-n junction and conducting one in p^+, n^+ layer return the carriers into SCR.

2. Recombination of carriers in n layer.

After switching the load current flows through SAS in reversal direction and pull out (extract) carriers into p^+, n^+ layers. The process resembles to the extraction process in DSRD, but there are differences which will be discussed.

The current density in SAS is much (order of magnitude) larger than in DSRD. Therefore during the extraction process, when part of plasma is removed and space charge region appears the voltage drop on SAS sharply rises and leads to high energy losses. It is well known fact that if space charge region does not overlap the n-layer (the usual condition in SAS before fast rise triggering voltage applied), in general case only part of plasma may be extracted. Nearly complete extraction of plasma is possible only in DSRD due to the matched processes of plasma injection and extraction by the current. In SAS both processes are of different nature and can not be matched, so the complete extraction of plasma by current is impossible.

Actually for good efficiency the width of current pulse must be so short that during the current pulse only small fraction of generated plasma is removed from the SAS volume.

It follows that the main way of plasma dispersal is recombination and for high PRF the lifetime of carriers must be short.

The time (τ_r) needed for concentration to drop from n_m down to n_0 level is

$$\tau_r = \tau_p \ln \frac{n_m}{n_0}, \tag{6.67}$$

where τ_p is life time of the minority carriers.

The $\tau_p < 1$ μs is needed to get $\tau_r \approx 20$ μs and $f_p \approx 50$ KHz (the case of $n_m \approx 10^{16}$ cm^{-3}, $n_0 \approx 10^6$ cm^{-3}). It is common knowledge that component of leakage current in SCR due to thermal carriers generation j_l is inversely proportional to the life time.

$$j_l \approx \frac{q n_i W_{SCR}}{\tau_p}, \tag{6.68}$$

where n_i - intrinsic concentration ~10^{10} cm^{-3} in Silicon at 20°C temperature, W_{SCR} SCR width.

In the case of $\tau_p \approx 1$ μs and $W_{SCR} \approx 10^{-2}$ cm (6.68) yields $j_l \approx 10^{-5}$ A/cm^2 and for carrier concentration in space charge region (n_0) we have:

$$n_0 \approx \frac{j}{1 V_S} \approx 10^7 cm^{-3}$$

very high value, more than $n_0 \approx 10^6$ cm^{-3} supposed before.

Nevertheless it is possible to switch on SAS with good efficiency for the case of large n_0 if applied triggering dU/dt voltage rate is increased up to $2 \cdot 10^{12}$ V/sec for each p-n junction in SAS.

In this case the estimation from (6.67) should be corrected by use of a new value for n_0, but the correction is minor.

It may be noted that n_i in (6.68) is increased exponentially with temperature and as has been mentioned that dependence is the main heat limitation of SAS.

Experiments showed that increase of dU/dt voltage rate up to $4 \cdot 10^{12}$ V/sec forces efficient SAS switching on even at PRF 200 KHz in burst mode at average PRF 10 kHz. SAS used in the experiment has $\tau_p > 2$ μs.

Two pulse experiments demonstrated that PRF as high as 500 KHz and more is possible in SAS with short life time of carriers. Of course such high PRF is possible only in burst mode. Average PRF is limited by heating, that additionally increases the leakage current.

It should be remarked that we discovered superfast delayed switching on in Silicon high voltage diodes with very short life time of carriers (less than 10 ns). Such short life times have been seen in diodes subjected to irradiation with light ions having energies > 5 megaelectronvolts. PRF of several MHz is expected for such structures. Their leakage current is far smaller than expected from (6.68), the fact may be explained by strong trapping of carriers on deep levels which were generated by irradiation. We consider such structures to be very effective as SAS, but their investigation are still at very early stage.

6.5 Circuits design for different shape of pulse generation

6.5.1 General consideration

The new devices are opening (DSRD) and closing (SAS) switches. Theirs main features as was mentioned are:

a) they have one stable state (off state) in which they can be infinitely long time, other state (on state) is relatively short

b) they are two electrode devices and triggering pulse is applied to the same electrodes through which load current passes.

Due to these features, the devices are most efficient when used in pulse compression circuits with bell-like shapes of pulses the main of which have been considered before in parts 1.2 (Fig. 1.5) and 6.2 (Fig. 6.1). Nevertheless the devices may be used for generation and shaping of large variety of different forms. As a rule in these cases their (device's) performance is degraded and/or some additional limitations appear.

Let us consider step like pulses. Some examples have been given in part 1.2.

Combined thyristor-diode closing switch (TDCS) shown in Fig. 1.7. TDCS emulates power closing switch with turn on time as fast as DSRD's turn off time. The method has disadvantage - all energy stored in additional capacitor C_p is lost. The lost energy fraction is equal to the ratio of energy stored in C_p and total energy stored in pulse forming network (PFN) (in the case of Fig. 1.7 $C_1 + C_2 + C_3$).

DSRD may be used to shape the front of pulse generated by other means (Fig. 1.6). Due to limitation for the length of pumping current, the separating inductor L_S can not be very large. The inductor shunts the line after opening of DSRD and distorts the flat part of the pulse (bends it down).

It may be shown that the decay of pulse due to pumping circuit (τ_d) is

$$\tau_d \approx \frac{U_+ \tau_+^2}{U_p \tau_{in}}, \qquad (6.69)$$

where U_+ - the voltage of pumping source, τ_+ - pumping pulse length, U_p - shaped pulse amplitude, τ_{in} - pulse front before shaping. In the case of $U_+ = U_p$, $\tau_+ \approx 200$ ns, $\tau_{in} = 40$ ns (6.69) yields $\tau_d \approx 1$ μs, that is the length of flat part of the shaped pulse is limited by the value ~1 μs.

In the case of SAS (see Fig. 1.9) the rate of voltage rise must be high > 10^{12} V/s, the front to be shaped less than 3 ns and thus, the pulse to be shaped must be generated by one of the DSRD circuits described above.

After switching on, the plasma generated in the diode is extracted by the current, SCR appears and voltage drop on SCR distorts the flat part of the pulse. The time of plasma extraction in diodes is less than 10 ns, that is the pulse decay is limited by the same value.

In sections 5.5, 5.6 it have been shown that in transistor and thyristor like structures after switching on, the plasma extraction is suppressed due to regeneration. SAS based on thyristors may be in "on" state indefinitely and they are the best choice for such application.

Exponential pulses with short front and long decay may be generated by symmetrical LC circuit shown in Fig. 1.5. In the case the decay time $\tau_d \approx L/R_l$ should be larger than turn off time of DSRD. It should be remembered that in the case of long decay time, considerable part of energy will return to the capacitors C_1, C_2 in accordance with the expression (6.28). The pulse may be shaped farther by use of SAS (see Fig. 1.9). In this case the use of SAS on thyristors is preferable as well.

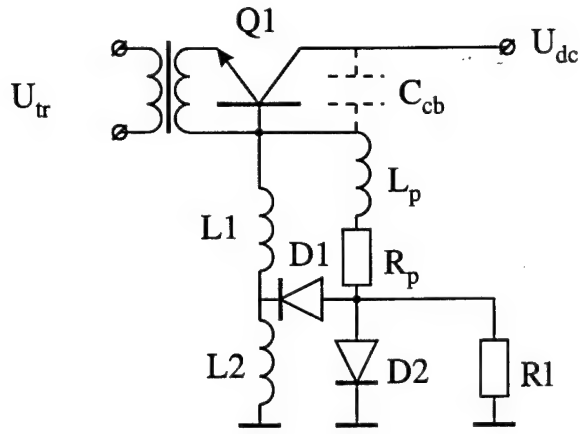

Q1
U_{dc}
U_{tr}
C_{cb}
L_p
L1
D1
R_p
L2
D2
R1

Fig.6.8

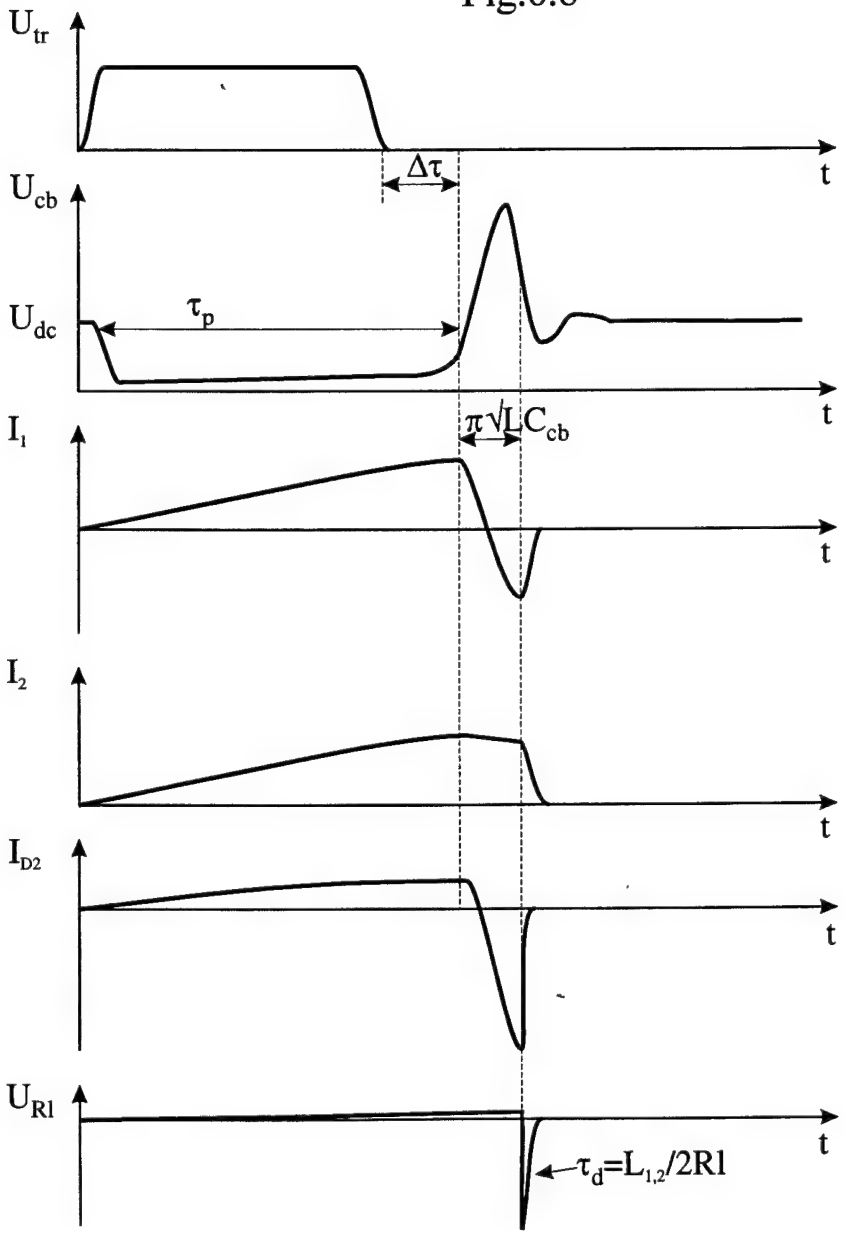

U_{tr}
$\Delta\tau$
t
U_{cb}
U_{dc}
τ_p
t
$\pi\sqrt{LC_{cb}}$
I_1
t
I_2
t
I_{D2}
t
U_{R1}
t
$\tau_d=L_{1,2}/2R1$

Fig.6.9

Rectangular pulses with short leading and trailing edges may be effectively generated by the same as above circuit (Fig. 1.5). The inductors L_1 and L_2 must be substituted by transmission lines T_1, T_2 or PFNs with impedances ρ_1, ρ_2.

The operation of the circuit is very close to circuit with inductors. The energy from capacitors C_1, C_2 is transferred to the magnetic field in lines. When DSRD breaks the current, the lines discharge their energy into the load. If the line impedance is matched with the load $R_l = \rho_1/2 = \rho_2/2$, the circuit generates the rectangular pulse with flat summit and length $2\tau_l$, where τ_l is the delay time of the line. The load current I_l is

$$I_l = \frac{U_0\sqrt{C_1}}{\sqrt{\tau_l \rho_1}} = I_1 = I_2 \qquad (6.70)$$

where U_0 - is the voltage on capacitor C_1.

It should be noted that during the pulse current flowing the capacitors C_1 and C_2 are recharged and part of the energy stored in lines is lost in accordance with (6.28). Therefore only short pulses ($2\tau_L >> T$, where T is the half period of LC circuit) could be generated with good efficiency.

For long pulses (> 100 ns) the circuit with TDCS (Fig. 1.7) has better efficiency.

6.5.2 High frequency circuits

DSRT (transistors) may be very effectively used for generation of pulses at very high PRF. The circuit is shown in fig. 6.8.

When triggering pulse is applied to the base-cathod electrodes, in accordance with the process, described in the part 4.4, DSRT (Q_1) closes. The current in the inductors L_1, L_2 rises (see Fig. 6.9). The diode D_1 prevents the inductor L_2 to be shunted by the diode D_2 (DSRD). The current I_p of the circuit R_pL_p pumps D_2.

Some time latter after the end of the triggering pulse, DSRT quickly opens and breaks conductivity current. The inductor L_2 current (I_2) passes through the diode D_1 into the DSRD (D_2) in reverse direction. Due to the low resistance of the diodes D_1, D_2, which are in the highly conducting phase, the I_2 current decay (τ_{dec}) is very slow $\tau_{dec} \approx L_2/R_f$, where R_f is the low resistance of D_1, D_2.

The inductor L_1 current I_1 charges the capacitance of DSRT collector C_{cb}. Then the current I_1 changes direction and passes through D_1 into the DSRD D_2 increasing the diode current I_D. At the moment of maximum of I_1 the total DSRD current I_D is two times more than I_2.

At the moment when I_D reaches the maximum value, DSRD breaks current and the energy stored in the inductors L_1, L_2 is transferred into the load. The front of the pulse is equal the diode's turn off time, the pulse decay ($\tau_d = L_1/2R_l$ if $L_1 = L_2$) as in the case of symmetrical circuit shown in fig.6.1.

After the end of the load current, the circuit is ready for the next cycle. The maximum repetition rate is limited by the time needed to store energy in the inductors ≤ 100 ns plus the time of half period of oscillation of the circuit consisting of inductor L_1 and DSRT collector capacitance < 10 ns.

Experiments demonstrated PRF more than 5 MHz.

6.5.3 Wave shaping line

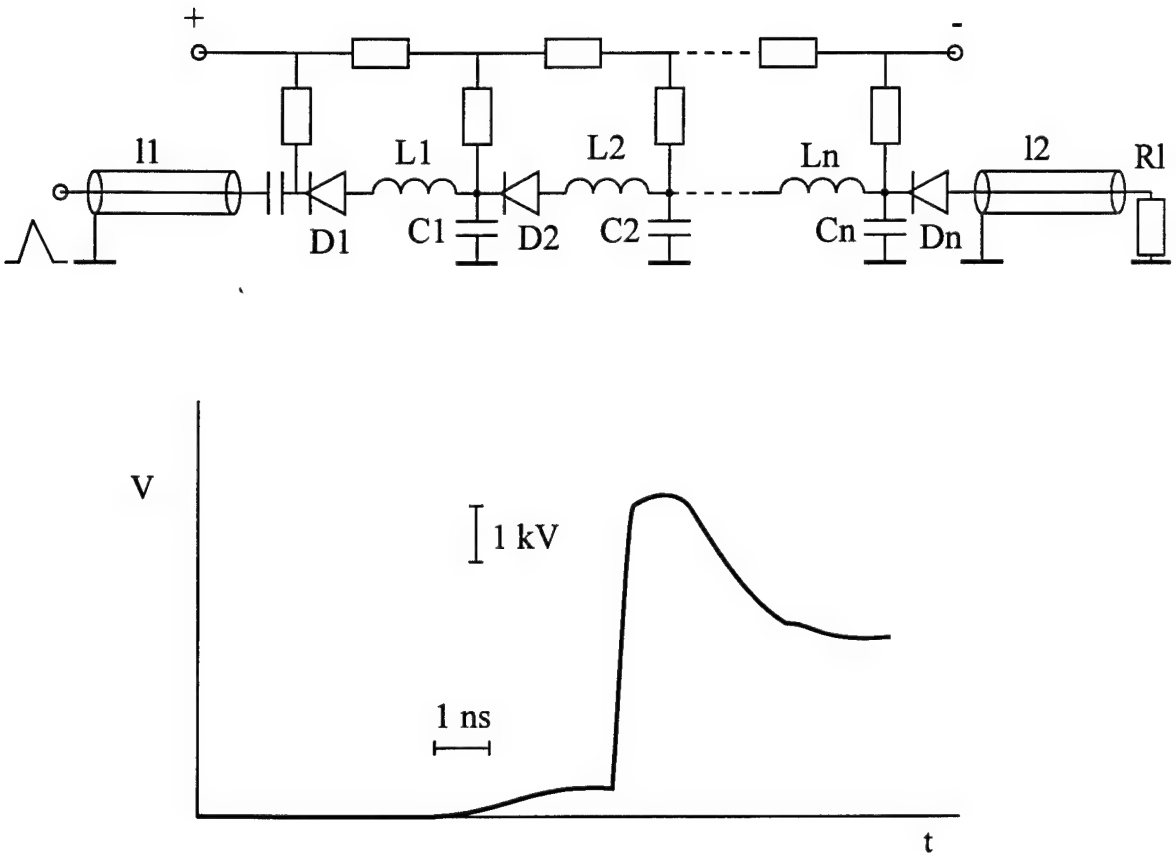


Fig.6.10

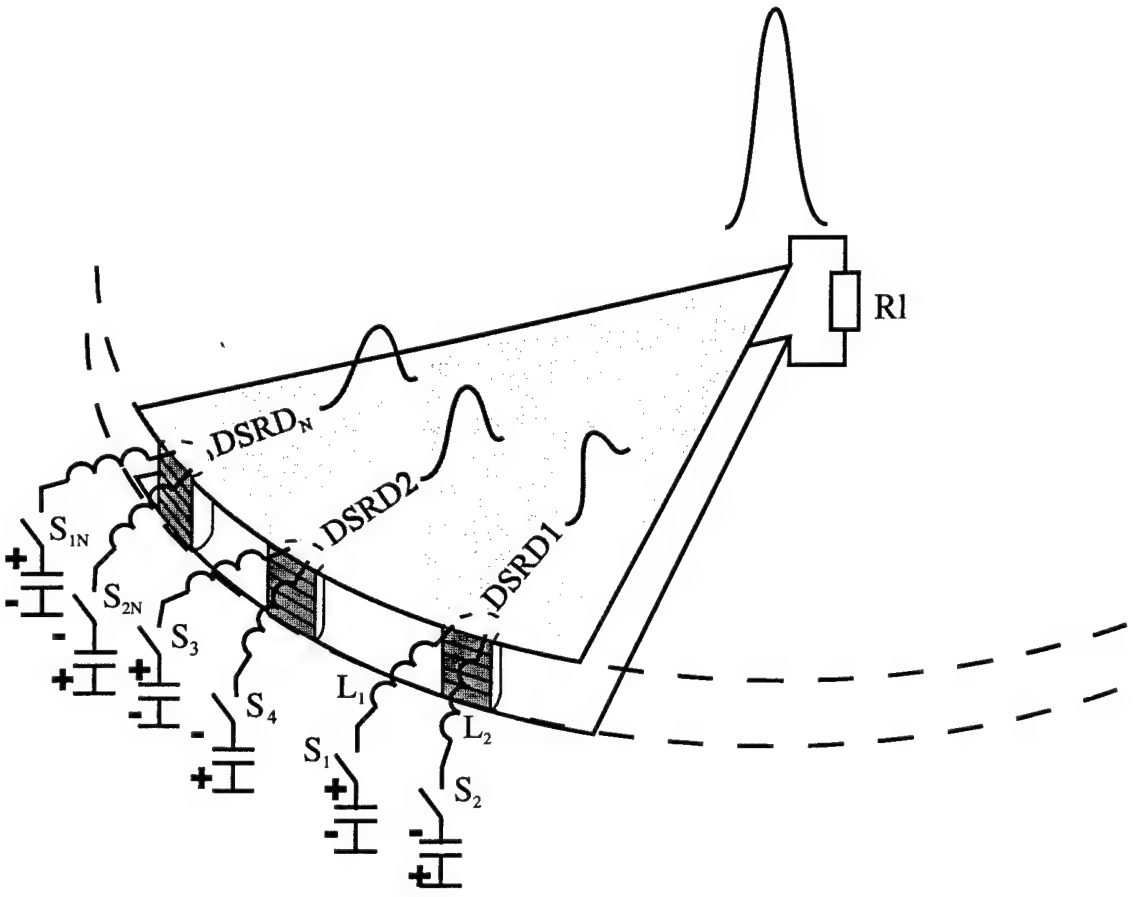


Fig.6.11

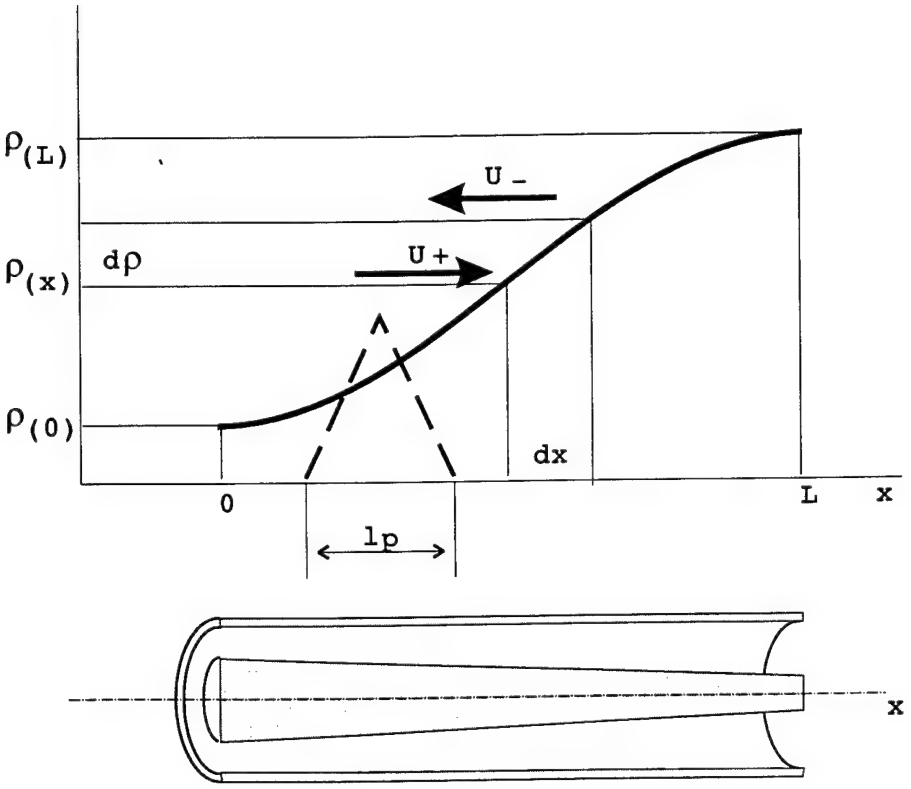


Fig.6.12

There is a method to generate efficiently short pulses, based on the fact that the SAS's are switched on like sequential elements of an transmission line on LC cells, matched with the input (l_1) and output (l_2) lines (fig.6.10). The constant bias voltage between the diodes is distributed using resistive networks. The voltage, charging capacitors C_n increases along the line. The pulse falling on the first SAS switches it on, and in the next cell of the line, a wave with a large amplitude is shaped. During the transition to the next cell due to the "cut-off of the upper frequency" effect which is known in discreet lines, the duration of the wave's front increases to a value which is close to the cell half period (cell's time constant) . Then the next SAS is switched on, and its amplitude again increases when the duration of the front is retained, etc.

In a line made of four SAS's, when switching three cells with a time constant of 0.5 ~ns and a line impedance of 50 ohms between each SAS pair, a pulse is shaped at the output which has an amplitude of 5 kV and a front duration of 0.3 ns (fig.6.10b). The peak power is 500 kW. The advantage of such a method consists of the fact that the energy for obtaining a sub-nanosecond pulse with a large amplitude is taken from the constant bias source, while the triggering generator shapes a wave with the amplitude necessary for switching on only the first SAS.

6.6 Matching with loads

6.6.1 General consideration.

As was shown in the section 6.1, it is possible adjust the voltage of DSRD stack (by the number of p-n junctions) and the current (by the area of p-n junction) to match the pulser with the load. The most used 50 Ohm load need, for example, 50 kV stack at 1 kA. Such stack consists of many (~50) p-n junctions, is thick and its cooling is severely limited and so is average power.

The same peak power due to high stability of DSRD circuits could be achieved when all DSRD work in parallel. In this case the cooling condition is many times (for the example - 50 times) better, but the pulser output impedance is very low 0.02 Ohm (1 kV, 50 kA). It should be noted that generally DSRD pulser has, as well, simpler design for lower impedance output.

In the cases when load impedance is fixed and can not be adjusted to pulser output, the impedance transformer (matching circuit) must be used (see Fig. 6.11).

To match pulser and load for sub and nanosecond pulsers only two methods are possible.

1. Line transformer based on transmission lines with variable wave impedances along its length.

2. Line transformer based on section of transmission lines with constant wave impedances.

The combination of these methods is possible.

6.6.2 Line transformer based on transmission lines with variable wave impedances along its length.

Let us consider line transformer of length L (Fig.6.12) with variable impedance $\rho(x)$. Reflection ratio $dK(x)$ from small part of line dx with impedance change $d\rho$ is

$$dK(x) = \frac{1}{2\rho(x)}\frac{d\rho(x)}{dx}dx = \frac{1}{2}\frac{d\ln\rho(x)}{dx}dx = N(x)dx \qquad (6.71)$$

$N(x) = \frac{1}{2}\frac{d\ln\rho(x)}{dx}$ is called local reflection function.

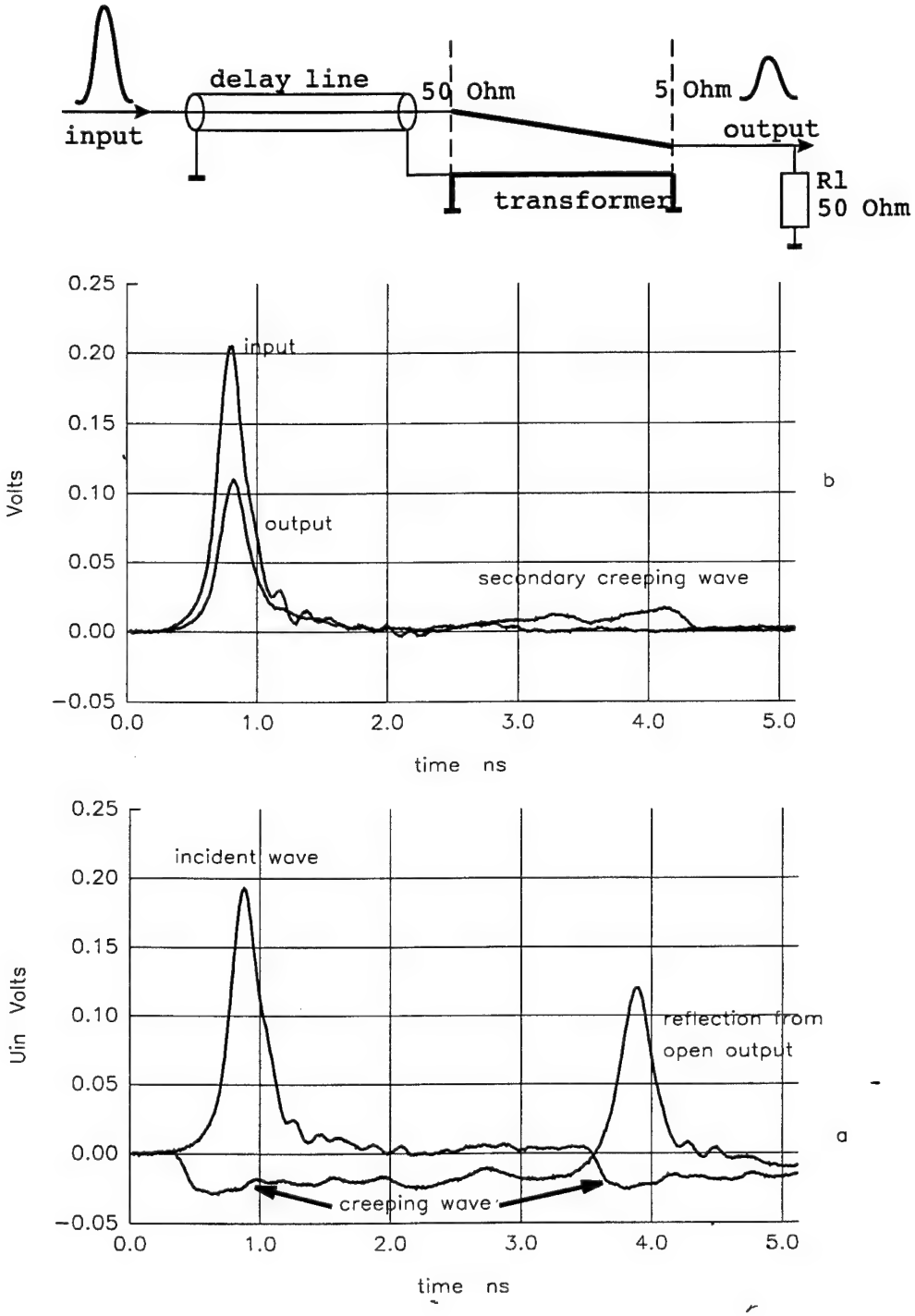

delay line
50 Ohm
5 Ohm
input
output
transformer
Rl
50 Ohm
0.25
0.20
0.15
0.10
0.05
0.00
−0.05
Volts
input
output
secondary creeping wave
b
0.0
1.0
2.0
3.0
4.0
5.0
time ns
Uin Volts
incident wave
reflection from
open output
creeping wave
a
time ns

Fig.6.13

It is evident that the next condition is valid

$$\frac{1}{2}\ln\frac{\rho(L)}{\rho(0)} = \int_0^L N(x)dx \qquad (6.72)$$

Increment of reflected wave (dU_) on distance dx is

$$dU_- = U_+N(x)dx \qquad (6.73)$$

were U_+ is incident wave.

In general, incident wave attenuates due to reflection and it is impossible to evaluate reflected wave from (6.73). It is possible only in the case of weak reflection and attenuation .

Assuming weak attenuation, from (6.73) we have

$$U_- = \int_0^{l_p} U_+N(x)dx \qquad (6.74)$$

where l_p is the incident pulse length in line.

When N(x) is constant, line impedance varies exponentially with distance (N(x) > 0 increases, N(x) < 0 decreases). Let us consider the simplest case of triangle incident wave with amplitude U_m and width l_p. For the case of N(x) constant, and $l_p < L$ from (6.74) we have, that for $t > \frac{l_p}{C}$ reflected wave amplitude (U_r) is constant

$$U_r = \frac{U_m N l_p}{2} \qquad (6.75)$$

From (6.72) - (6.75) we can evaluate the loss of energy due to reflection (Q_)

$$Q_- \approx \frac{U^2 L}{\rho(0)} \approx Q_+ \frac{l_p\left(\ln\frac{\rho(L)}{\rho(0)}\right)^2}{4L} \qquad (6.76)$$

where Q_+ - incident wave energy

L - line length

$\rho(0)$, $\rho(L)$ - wave impedances at the line ends.

The expression (6.75) is valid in the case of weak reflection, i.e $Q_- << Q_+$.

From (6.76) we have that reflection losses are small when the line length is much more than the pulse length and transformation ratio (Ku) $Ku = \frac{U(L)}{U(0)} = \sqrt{\frac{\rho(L)}{\rho(0)}}$ is not high.

We have checked validity of (6.74) - (6.76) by computer simulation and by experiment for $\rho(L) = 5$ 0hm, $\rho(0) = 50$ Ohm, FWHM $\tau_p = \frac{l_p}{C} = 0.25$ ns, $\tau_1 = \frac{L}{C} = 1.75$ ns. The line transformer end with 5 Ohm output was opened (output load resistance is 50 Ohm (Fig.6.13)). The results of time - domain reflectometry are shown at Fig.6.13a,b . Nearly constant (in accordance with (6.75)) reflected creeping wave is clearly seen. Creeping wave amplitude is ≈ 8% of initial incident wave, that value matches the equation (6.75).

Output pulse at not matched (50 Ohm load) end for no losses case must be ≈ 0.575 of input (0.316 for matched case). From (6.76) we have reflection energy loss $Q_- \approx 0.18\ Q_+$ and corresponding 90% amplitude decrease for output pulse. In accordance with this value output amplitude is ≈ 0.517 of input. Experiment (Fig.6.13) gives 0.54 trifle over calculated 0.517. The difference is due to incident wave attenuation, clearly seen from Fig.6.13 (that has not been taken into an account in (6.76)).

The creeping reflected wave in turn is reflected as well and gives rise to the tail incident wave seen after output pulse (Fig.6.13b).

From (6.76) we have that efficient line transformer with high transformation ratio must be long.

Our computer modeling showed that efficiency depends only slightly on the type of function N(x). The difference for all tested functions N(x) (N(x)= x(L-x), N(x)= th(x) , N(x) = const, $N(x)=\sin\frac{x\pi}{L}$) was inside several percents.

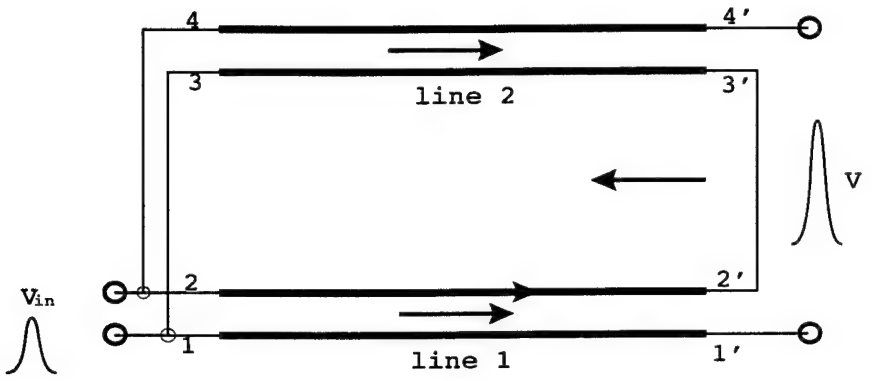
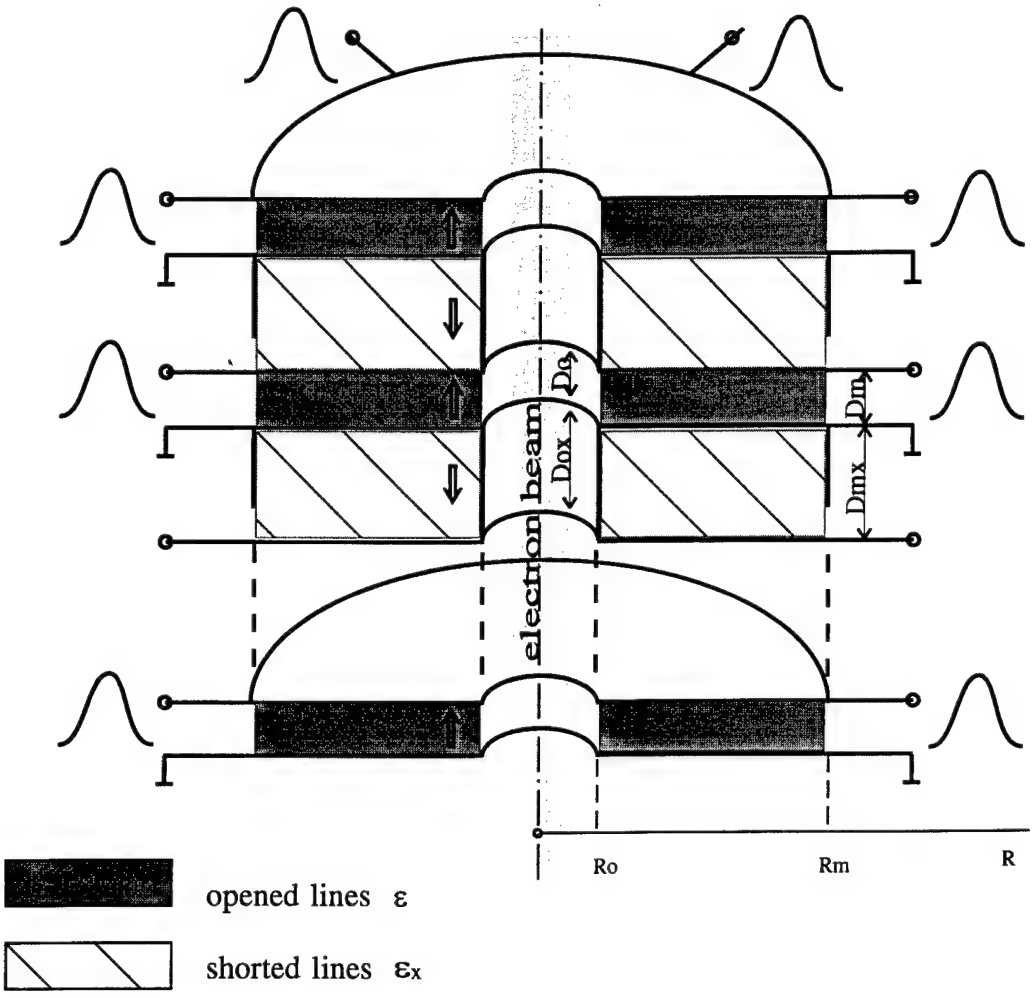
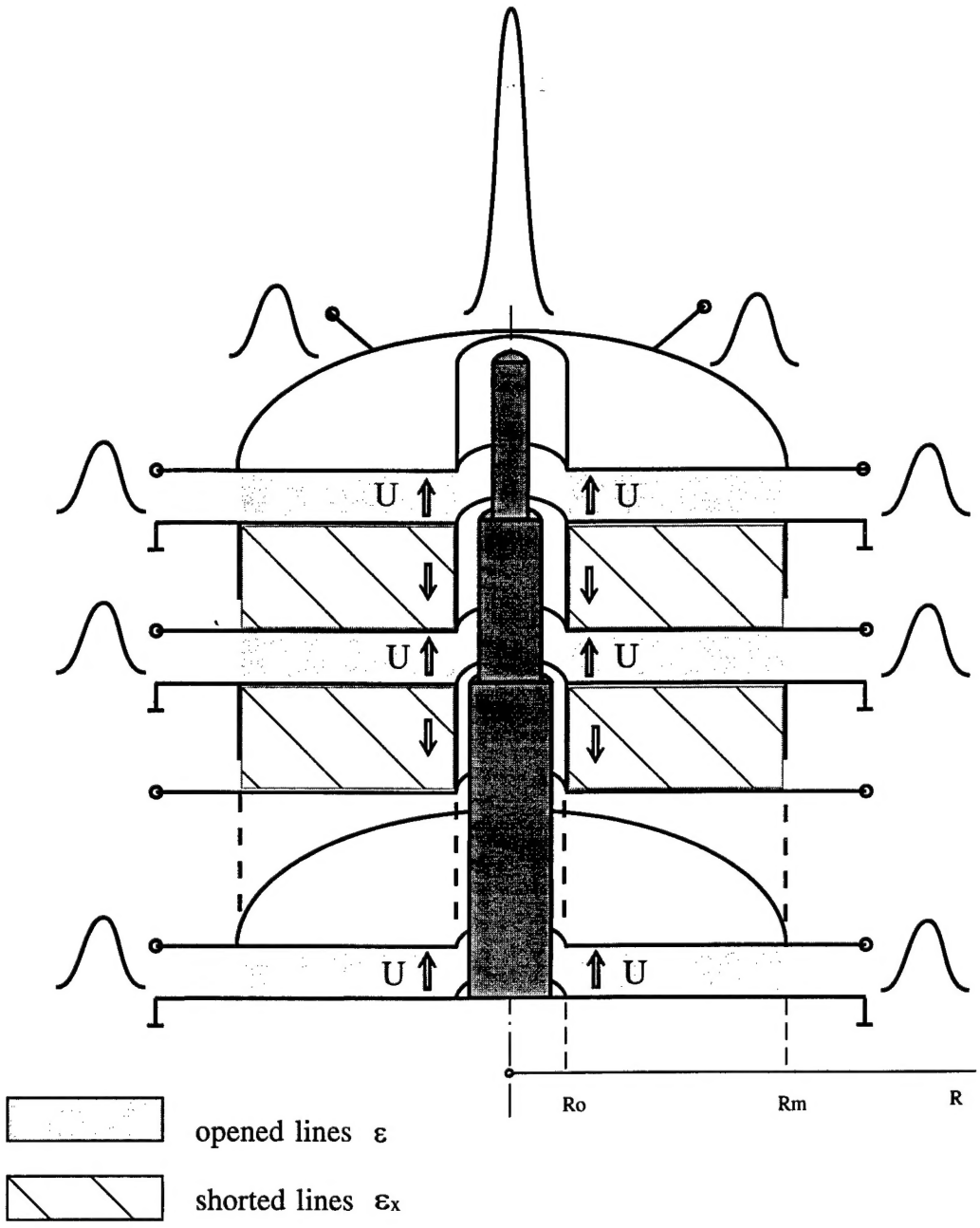


Fig.6.14



Radial line ironless acselerator

Fig.6.15



Radial line nanosecond summator

Fig.6.16

6.6.3 . Transformer based on section of transmission lines with constant wave impedances.

Transformer on sections of transmission lines is shown in Fig.6.14. Two line inputs are connected in parallel and their outputs in series. The input impedance (ρ_{in}) is ρ_0 /2, output (ρ_{out}) is $2\rho_0$, where ρ_0 is line impedance. For n lines $\rho_{in}=\rho_0/n$, $\rho_{out}=n\rho_0$ and transformation ratio Ku = n. The transformer can work in reverse way.

It may be seen from Fig.6.14 that in addition to two main lines (1,1 ' 2,2') and (3,3 ' ;4,4') parasite line (1,1' ; 3,3') is present. When incident wave achieves end (1' , 2') of the line 1, parasite wave, traveling to its shorted end (1,3), is exited in the parasite line. The ratio of powers of parasite to main waves is determined by the next ratio.

$$\frac{P_{W_1}}{P_{W_2}} = \frac{\rho_2}{\rho_1} \qquad (6.77)$$

There exist the next techniques to increase wave impedance ρ_2 and decrease parasite wave.

a. To make transmission line - as thin, as possible. But high power short pulses require for good transmission rather thick cables .

b. To fill the space, surrounding the conductors, with high wave impedance (ρ_f) media

$$\rho_f = \sqrt{\frac{\mu_0\mu_f}{\varepsilon_0\varepsilon_f}} \qquad (6.78).$$

where μ_0 - permeability, ε_0 - dielectric permittivity of vacuum, μ_f, ε_f- their relative values for media.

The simplest way is to "dress" the cable line with ferrite rings. It is well known, that for ferrite relative value of μ_f may achieve hundreds, while ε_f is low: 3-4 . But due to magnetic domain inertia for nano and subnanosecond pulses the "effective" μ_f is much lower.

Our experiments with NN400 ferrite showed that for times less than 70 ps $\mu_f / \varepsilon_f < 1$ due to high value of ε_f and poor μ_f, then μ_f increases and 200 ps later ρ_f achieves 3 times of the vacuum value. It seems that for sub and nanosecond pulses this technique is possible, but not very effective.

c. To coil up the feeder (cable). This method is very effective with thin flexible feeders. For thick and rigid high voltage feeder the method is not very good in practice.

Two last techniques or even all three of them may be combined. The feeder coil may be wound around ferrite rode.

One can easily find, that number of possible unwanted parasite lines (m) increases with number of main lines n super linearly as

$$m = \frac{n(n-1)}{2} \qquad (6.79)$$

For each from this m parasite lines the parasite wave power relation (6.77) is valid. It should be noted, that for some of the unwanted lines of higher order the exciting voltage Vn may exceed the input voltage. For example the voltage for parasite line consisting of the uppermost and the lowest lines will be (n-1) times more then input voltage, so the power of parasite waves may be high .

Thus, transformers with high transformation ratio (n>3) become very complicated and not very effective.

Both types of transformers work in reverse direction and may be used to match low impedance output with load.

The transformer shown in Fig. 6.11 at the same time sums up the power of several pulsers and matches the impedances. The width of transformer's line may be increased to

accommodate larger number of pulsers. The limit is radial line with the load placed in the center of disk and the pulsers placed at the circumference.

Such radial line transformers may be in their turn assembled in a stack as shown in Fig. 6.15 to sum up the output voltage. The inner wire may be changed by electron beam to design an electron accelerator (see Fig.6.16).

In [6.1] such nanosecond solid state accelerator has been suggested and calculated at 1 MV, 100 kA, peak power 10^{11} W output and average power > 1 MW, with efficiency ≈ 70%.

References to the part 6.

[1] High power electronics IX Conference, 21-30 July, 1992, Russia, p. 303-304. I.V.Grekhov. A.f.Kardo-Sysoev, \ The approach to design of power nanosecond solid-state ironless accelerators and high voltage generators based on semiconductor devices. \

7 Conclusion

In recent times, the physics and engineering of the fast commutation of high electric powers using semiconductor devices are developing extremely rapidly. Even during the time of preparing this report new interesting results were obtained. Therefore we found it necessary in the conclusion to discuss briefly the degree of completion of the pattern for the physical processes, described in this report in the devices, the means and possibilities for improving the parameters of the devices, pulsers and the last achievements in this area.

It can be affirmed with confidence that the physical mechanism for the operation of drift step recovery devices with fast restoration is understood somewhat completely, and that, following only from the most general principles, it is not complicated to calculate the design parameters for an ideal semiconductor structure (synthesis), that is, structures whose basic operational parameters (working voltage, duration of high conductivity stage, break-off current, current break-off time) will have the best physically possible value. It is simple enough to solve the analysis problem as well: following from the design parameters, to calculate the operational parameters of a specific structure, at least diode and transistor. The main unsolved problem is technological one -- how to make a structure which is close to ideal.

Very promising approach - direct wafer bonding - was successfully tested at Ioffe PTI to manufacture DSRD and SAS, but a lot of additional work and investigation should be done before wide practical application of this technology in the pulser design.

Physics of the processes in devices with delayed ionization is immeasurably more complicated, and it has obviously not been studied sufficiently. The existence in ionization phenomena of different types of instabilities and also the lack of reliable data on the ionization processes with the participation of deep levels (states) in the forbidden gap also impede synthesis and analysis with an accuracy which is acceptable for practical purposes. Such device parameters as the switching time and the residual voltage are almost not being subjected to calculation at the present time, while the designing of SAS and pulse shapers is mainly going on in an empirical fashion.

Let us recall that the maximum possible fast response for devices with fast restoration is determined by the maximum speed of restoration of the voltage, which is for a p-n junction approximately equal to ~2 10^{12}V/s, which, when the front duration equals 2 ns, correspond s to a working voltage of 2 kV and a degree of doping level of n-layer of 10^{14} cm^{-3}. The dU/dt

value may be increased proportionally ot a number of p-n-junctions assembled in a stack. The voltage of a stack may be as high as hundreds kV. The density of the break-off current is >160 A/cm^2 and the current itself, which is limited by skinning, equals approximately 10kA when the diameter of the structure is 10 cm. Since DSR devices are easily synchronized, it is possible to increase the switched voltage and the current by means of the series and parallel switching on of a large number of devices.

In this case fundamental limitations on increasing the power are not seen.

For devices with delayed ionization, the threshold for the minimum switching time (approximately $10^{-12}s$) , which is the time between two sequential acts of ionization when there is the saturated value for the ionization coefficient, is clear. Instabilities can considerably worsen the commutation parameters, and they also do not make it possible to determine the prospects for increasing the switched power in the case of the series and parallel switching on of the devices. Nevertheless, the technology for the generation of sub-nanosecond pulses is developing quickly.

The recent achievements in the technology for shaping pulses, obtained by the authors' are the following: for a triangular pulse, a front duration of less than 0.1 ns when the amplitude is more than 20 kV on a load of 50 Ohms; that is, a pulse power of more than 8 MW at pulse duration 1 ns. For a bell-shaped pulse with a half width of approximately 2 ns and rise time 0.7ns, the amplitude is more than 90 kV on a load of 100 ohms; that is, a power of more than 80 MW and an efficiency of approximately 80%.

These parameters are unique, but in the near future they will also be significantly improved.